

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 06:59:35 ; Search time 59 Seconds
(without alignments)
1743.174 Million cell updates/sec

Title: US-09-829-275-1

Perfect score: 1881

Sequence: 1 NM5GQKRLMWAGTGGHV.....RVANEVSRVAFLEHHHHH 364

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 100 summaries

Database : A_Geneseq_29Jan04.*

1: Geneseq1980s.*

2: Geneseq1990s.*

3: Geneseq2000s.*

4: Geneseq2001s.*

5: Geneseq2002s.*

6: Geneseq2003as.*

7: Geneseq2003bs.*

8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|--------------------|
| 1 | 1841 | 97.9 | 364 | 5 | Aau99816 Escherich |
| 2 | 1803 | 95.9 | 355 | 4 | Aau34438 E. coli c |
| 3 | 1803 | 95.9 | 355 | 4 | Ag98406 Escherich |
| 4 | 1803 | 95.9 | 355 | 6 | Abu28491 Protein e |
| 5 | 1687 | 89.7 | 355 | 6 | Abu47265 Protein e |
| 6 | 1677 | 89.2 | 355 | 4 | Aau38485 Salmonell |
| 7 | 1677 | 89.2 | 355 | 6 | Abu48097 Protein e |
| 8 | 1604 | 85.3 | 348 | 6 | Abu31657 Protein e |
| 9 | 1549.5 | 82.4 | 348 | 6 | Abu45116 Protein e |
| 10 | 1471 | 78.2 | 356 | 6 | Abu50002 Protein e |
| 11 | 1431 | 76.1 | 356 | 6 | Abm68867 Phototrab |
| 12 | 1345.5 | 71.5 | 360 | 6 | Abu40887 Protein e |
| 13 | 1323 | 70.3 | 281 | 6 | Abu27914 Protein e |
| 14 | 1154 | 61.4 | 354 | 6 | Abu49532 Protein e |
| 15 | 1123 | 59.7 | 354 | 6 | Abu38975 Protein e |
| 16 | 1064.5 | 56.6 | 351 | 4 | Aau35596 Haemophil |
| 17 | 1064.5 | 56.6 | 351 | 6 | Abu30498 Protein e |
| 18 | 1060 | 56.4 | 351 | 5 | Aau99817 Haemophil |
| 19 | 871 | 46.3 | 367 | 6 | Abu22301 Protein e |
| 20 | 849 | 45.1 | 372 | 6 | Abu21174 Protein e |
| 21 | 791 | 42.1 | 355 | 6 | Abu23450 Protein e |
| 22 | 772.5 | 41.1 | 357 | 6 | Abu38137 Protein e |
| 23 | 764.5 | 40.6 | 355 | 3 | Aay74403 Neisseria |
| 24 | 759.5 | 40.4 | 355 | 3 | Aay74401 Neisseria |
| 25 | 759.5 | 40.4 | 355 | 6 | Abp80549 N. gonorr |

| | | | | | |
|----|-------|------|-----|---|--------------------|
| 26 | 759.5 | 40.4 | 355 | 6 | ABU37201 Protein e |
| 27 | 719.5 | 38.3 | 357 | 4 | AAU36416 Pseudomon |
| 28 | 719.5 | 38.3 | 357 | 6 | ABU38727 Protein e |
| 29 | 717 | 38.1 | 356 | 6 | ABU41785 Protein e |
| 30 | 696.5 | 37.0 | 367 | 6 | ABU35148 Protein e |
| 31 | 684 | 36.4 | 375 | 6 | ABU40398 Protein e |
| 32 | 680.5 | 36.2 | 366 | 6 | ADA33749 Acinetoba |
| 33 | 672.5 | 35.8 | 365 | 6 | ABU16742 Protein e |
| 34 | 630.5 | 33.5 | 335 | 3 | AAY74402 Neisseria |
| 35 | 535 | 28.4 | 408 | 6 | ABU34101 Protein e |
| 36 | 534 | 28.4 | 250 | 6 | ABU19873 Protein e |
| 37 | 510.5 | 27.1 | 407 | 6 | ABU35864 Protein e |
| 38 | 506 | 26.9 | 410 | 6 | ABU15855 Mycobacte |
| 39 | 506 | 26.9 | 410 | 6 | ABU15855 Mycobacte |
| 40 | 506 | 26.9 | 410 | 6 | ABU36727 Protein e |
| 41 | 506 | 26.9 | 410 | 6 | ABU34407 Protein e |
| 42 | 488.5 | 26.0 | 359 | 6 | ABU26097 Protein e |
| 43 | 487 | 25.9 | 372 | 4 | AAG92109 C glutami |
| 44 | 479 | 25.5 | 380 | 6 | ABU20900 Protein e |
| 45 | 477.5 | 25.4 | 409 | 5 | AAU99823 Mycobacte |
| 46 | 470.5 | 25.0 | 409 | 6 | ABU25498 Protein e |
| 47 | 460 | 24.5 | 383 | 6 | ABU17984 Protein e |
| 48 | 451.5 | 24.0 | 386 | 4 | AAU60464 Propionib |
| 49 | 451.5 | 24.0 | 386 | 6 | ABM56983 Propionib |
| 50 | 435 | 23.1 | 357 | 5 | ABBS4935 Lactococc |
| 51 | 431.5 | 22.9 | 393 | 5 | ABP66056 Bifidobac |
| 52 | 423.5 | 22.5 | 366 | 6 | ABU18309 Protein e |
| 53 | 422 | 22.4 | 363 | 5 | AB848474 Listeria |
| 54 | 422 | 22.4 | 363 | 6 | ABU32852 Protein e |
| 55 | 420 | 22.3 | 362 | 6 | ABU29950 Protein e |
| 56 | 420 | 22.3 | 363 | 7 | ADC97130 E. faeciu |
| 57 | 411.5 | 21.9 | 363 | 6 | ABU14640 Protein e |
| 58 | 406.5 | 21.6 | 387 | 6 | ADB11850 Alloioccc |
| 59 | 402.5 | 21.4 | 368 | 4 | AAU35044 Enterococ |
| 60 | 401.5 | 21.3 | 363 | 4 | AAU33370 Enterococ |
| 61 | 393.5 | 20.9 | 382 | 5 | AAU99822 Bacillus |
| 62 | 370 | 19.7 | 362 | 5 | AAU99818 Enterococ |
| 63 | 368.5 | 19.6 | 359 | 5 | AAU99819 Enterococ |
| 64 | 335.5 | 17.8 | 354 | 6 | ABU24880 Protein e |
| 65 | 332 | 17.7 | 431 | 5 | ABB91578 Herbicida |
| 66 | 328.5 | 17.5 | 359 | 6 | ABU23929 Protein e |
| 67 | 315.5 | 16.8 | 384 | 6 | ABU48604 Protein e |
| 68 | 313.5 | 16.7 | 385 | 5 | AAU99821 Rickettsi |
| 69 | 310 | 16.5 | 326 | 2 | AAU37382 Protein 1 |
| 70 | 307 | 16.3 | 352 | 7 | ADD43922 Chlamydia |
| 71 | 303 | 16.1 | 353 | 6 | ABU27160 Protein e |
| 72 | 284.5 | 15.1 | 357 | 2 | AAU35559 Chlamydia |
| 73 | 284.5 | 15.1 | 357 | 5 | ABB90579 Chlamydia |
| 74 | 284.5 | 15.1 | 357 | 6 | ABU27032 Protein e |
| 75 | 279 | 14.8 | 353 | 5 | AAU99820 Streptoco |
| 76 | 277 | 14.7 | 353 | 4 | AAU35971 Helicobac |
| 77 | 275.5 | 14.6 | 352 | 4 | AAU38041 Streptoco |
| 78 | 275 | 14.6 | 353 | 6 | ABU30968 Protein e |
| 79 | 273.5 | 14.5 | 352 | 2 | AAW51348 Streptoco |
| 80 | 273 | 14.5 | 356 | 4 | AAU37021 Staphyloc |
| 81 | 273 | 14.5 | 356 | 4 | AAU37260 Staphyloc |
| 82 | 273 | 14.5 | 356 | 6 | ABU42410 Protein e |
| 83 | 273 | 14.5 | 356 | 6 | ABM73029 Staphyloc |
| 84 | 272.5 | 14.5 | 352 | 4 | AAU37806 Streptoco |
| 85 | 272.5 | 14.5 | 352 | 6 | ABU45942 Protein e |
| 86 | 269.5 | 14.3 | 352 | 6 | ABU01073 S. pneumo |
| 87 | 263 | 14.0 | 342 | 6 | ABU26566 Protein e |
| 88 | 261 | 13.9 | 357 | 6 | ABU42647 Protein e |
| 89 | 261 | 13.9 | 366 | 5 | ABP38353 Staphyloc |
| 90 | 260 | 13.8 | 357 | 4 | AAU81658 S. epider |
| 91 | 259.5 | 13.8 | 363 | 6 | ABU19335 Protein e |
| 92 | 250 | 13.3 | 360 | 5 | ABP29983 Streptoco |
| 93 | 250 | 13.3 | 360 | 6 | ABU46729 Protein e |
| 94 | 250 | 13.3 | 367 | 5 | ABP25519 Streptoco |
| 95 | 247 | 13.1 | 358 | 5 | ABP25518 Streptoco |
| 96 | 224 | 11.3 | 313 | 4 | AAU82061 S. epider |
| 97 | 213.5 | 11.4 | 361 | 6 | ABU44117 Protein e |
| 98 | 160.5 | 8.5 | 186 | 6 | ABU43476 Protein e |

99 154 8.2 104 5 ABP11313 Human ORF
100 150 8.0 554 4 AAG80912 MGD synth

ALIGNMENTS

RESULT 1

AAU99816
ID AAU99816 standard; protein; 364 AA.

XX AC AAU99816;

XX 07-OCT-2002 (first entry)

XX Escherichia coli Membrane associated UDP-glycosyltransferase MurG.

KW MurG; membrane associated UDP-glycosyltransferase; antibiotic;
KW antimicrobial; modulator of glycosyltransferase activity; drug design;
KW UDP-glycosyltransferase; directed drug design; random drug design;
KW grid-based drug design.

XX OS Escherichia coli.

XX Key Location/Qualifiers

FT Domain 7..163

FT /label= N-domain

FT Region 14..19

FT /label= G-loop_1

FT Misc-difference 75

FT /note= "Member of the N-domain hydrophobic patch"

FT Misc-difference 79

FT /note= "Member of the N-domain hydrophobic patch"

FT Misc-difference 82

FT /note= "Member of the N-domain hydrophobic patch"

FT Misc-difference 85

FT /note= "Member of the N-domain hydrophobic patch"

FT Region 102..109

FT /label= G-loop_2

FT Misc-difference 116

FT /note= "Member of the N-domain hydrophobic patch"

FT Domain 164..340

FT /label= C-domain

FT Region 190..196

FT /label= G-loop_3

FT Domain 341..357

FT /label= N-domain

XX WO200190301-A2.

XX 29-NOV-2001.

XX 09-APR-2001; 2001WO-US011500.

XX 17-MAY-2000; 2000US-0204930P.

XX (UYPR-) UNIV PRINCETON.

XX Walker S, Ha S;

XX WPI; 2002-171402/22.

XX Novel composition comprising crystalline form of MurG protein, a membrane

XX -associated UDP-glycosyltransferase involved in peptidoglycan

XX biosynthesis, for determining ability of chemical compound to bind MurG

XX protein.

XX Claim 14; Fig 3A; 222pp; English.

XX The invention describes a composition comprising a membrane associated

XX UDP-glycosyltransferase, MurG, preferably Escherichia coli protein in

XX crystalline form. A model of UDP-glycosyltransferase is useful in a

XX computer-assisted method of structure based drug design of bioactive

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC compounds, by providing and designing a chemical compound using the
CC model. The method further comprises synthesising the chemical compound,
CC and evaluating the bioactivity of the synthesised chemical compound. The
CC bioactivity is selected from inhibiting binding of a nucleotide donor
CC compound or an acceptor compound to the MurG protein, or inhibiting
CC association of the MurG protein to a membrane. Designing the chemical
CC compound involves computational screening of one or more databases of
CC chemical compounds in which the 3D structure of the compounds are known,
CC and interacting a compound identified by the screening step with the
CC model by computer. The step of designing involves directed drug design,
CC random drug design, or grid-based drug design. Designing involves
CC selecting compounds which are predicted to bind to or mimic the 3D
CC structure of the MurG protein. A modulator of glycosyltransferase is
CC useful as antibiotics or antimicrobial agents in animals, and
CC therapeutically or diagnostically in an animal. This is the amino acid
CC sequence of the Escherichia coli MurG protein crystallised in the
CC invention

XX Sequence 364 AA;

XX Query Match 97.9%; Score 1841; DB 5; Length 364;

XX Best Local Similarity 98.1%; Pred. No. 4e-175;

XX Matches 357; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 VMSSGQKELVMAGTGGHVPGLAVAHHLMAQGHVWLTADRMEDALVPKHGIEIDF 60

Db 1 VMSSGQKELVMAGTGGHVPGLAVAHHLMAQGHVWLTADRMEDALVPKHGIEIDF 60

QY 61 IRISGLRGKGIKALIAAPLIRFNARQARAIKAYKPDVVLGMGGYVSGPGLAASLGI 120

Db 61 IRISGLRGKGIKALIAAPLIRFNARQARAIKAYKPDVVLGMGGYVSGPGLAASLGI 120

QY 121 PVTLHEQNGIAGLTNKLARIATKNQAEPCAFNAEVVGNVPTDVLALPIPOORLAGR 180

Db 121 PVTLHEQNGIAGLTNKLARIATKNQAEPCAFNAEVVGNVPTDVLALPIPOORLAGR 180

QY 181 EGPVRLVVGSGQARILNQTMPQVAALGDSVLIWHQSGKGSQOSVQVAYAEAGQPQHK 240

Db 181 EGPVRLVVGSGQARILNQTMPQVAALGDSVLIWHQSGKGSQOSVQVAYAEAGQPQHK 240

QY 241 VTEPTDDMAAYANADVVVCSGALTVSEIAAAGLPALFVFPQHKDQOQYNNALPLEKAG 300

Db 241 VTEPTDDMAAYANADVVVCSGALTVSEIAAAGLPALFVFPQHKDQOQYNNALPLEKAG 300

QY 301 AAKIIEQPQLSVDVANTLAGMSRETLTMAERARAASIIPDATERVANEVSRVARALEHH 360

Db 301 AAKIIEQPQLSVDVANTLAGMSRETLTMAERARAASIIPDATERVANEVSRVARALEHH 360

QY 361 HHHH 364

Db 361 HHHH 364

RESULT 2

AAU34438

ID AAU34438 standard; protein; 355 AA.

XX AC AAU34438;

XX 14-FEB-2002 (first entry)

XX E. coli cellular proliferation protein #19.

XX Antisense; prokaryotic cellular proliferation protein; antibiotic;

XX antibacterial; drug design.

XX Escherichia coli.

XX WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US009180.

XX 21-MAR-2000; 2000US-0191078P.
PR 23-MAY-2000; 2000US-0206848P.
PR 26-MAY-2000; 2000US-0207727P.
PR 23-OCT-2000; 2000US-0242578P.
PR 27-NOV-2000; 2000US-0253625P.
PR 22-DEC-2000; 2000US-0257931P.
PR 16-FEB-2001; 2001US-0269308P.
XX (ELIT-) ELITRA PHARM INC.
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX WPI: 2001-611495/70.
DR N-PSDB; RAS52297.
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.
XX Example 3; SEQ ID NO 10031; 511pp; English.
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes
CC themselves and the encoded proteins. The prokaryotes used are Escherichia
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
CC antibodies capable of binding to the expressed proteins. The proteins can
CC be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous
CC nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence represents an essential prokaryotic
CC cellular proliferation protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX Sequence 355 AA;
SQ

Query Match 95.9%; Score 1803; DB 4; Length 355;
Best Local Similarity 99.2%; Pred. No. 2.5e-171;
Matches 352; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 2 MSGGCKLWVWAGTGGHVPGLAVAHLLMAQGWQVRLGTADRMEDLVPKHGIEDFI 61
DB 1 MSGGCKLWVWAGTGGHVPGLAVAHLLMAQGWQVRLGTADRMEDLVPKHGIEDFI 60
QY 62 RISGLRGKIKALIAAPLRIFNWQRAIMKAYKPDVVLGMGGYVSGPGLAWSLIGIP 121
DB 61 RISGLRGKIKALIAAPLRIFNWQRAIMKAYKPDVVLGMGGYVSGPGLAWSLIGIP 120
QY 122 VVLEHQNGIAGLTNKLWARIATKQWQRAEAFNAEVVGNPVRTDVLALPLPOORLAGRE 181
DB 121 VVLEHQNGIAGLTNKLWARIATKQWQRAEAFNAEVVGNPVRTDVLALPLPOORLAGRE 180
QY 182 GPVRLVVGSGQARIILNQTMFPQVAAKLGDSVTIWHQSGKGSQSQVEQAYAEAGQPHKV 241
DB 181 GPVRLVVGSGQARIILNQTMFPQVAAKLGDSVTIWHQSGKGSQSQVEQAYAEAGQPHKV 240
QY 242 TEFTDDMAAYADVVCRSGALTIVSEIAAGLIPALFVFPQKDRQQYNNALPLEKAGA 301
DB 241 TEFTDDMAAYADVVCRSGALTIVSEIAAGLIPALFVFPQKDRQQYNNALPLEKAGA 300
QY 302 AKIIEQPOLSDVAVANTLAGRSRETLTMAERARAASIPDATERVANEVSRVARA 356
DB 301 AKIIEQPOLSDVAVANTLAGRSRETLTMAERARAASIPDATERVANEVSRVARA 355

RESULT 3

AAAG98406
ID AAG98406 standard; protein; 355 AA.
XX
AC AAG98406;
DT 21-SEP-2001 (first entry)
XX Escherichia coli protein sequence SEQ ID NO:454.
XX Escherichia coli; identification; proliferation; microorganism;
KW antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
KW bacterial growth inhibition.
OS Escherichia coli.
XX WO200148209-A2.
XX 05-JUL-2001.
XX 19-DEC-2000; 2000WO-US034419.
XX 23-DEC-1999; 99US-0173005P.
XX (ELIT-) ELITRA PHARM INC.
XX Forsyth RA, Ohlsen KL, Zyskind JW;
XX WPI: 2001-457376/49.
DR N-PSDB; AAH81462.
XX Novel nucleic acids encoding proteins required for Escherichia coli
PT proliferation, useful for screening for antimicrobial agents.
XX Claim 19; Page 573; 596pp; English.
XX The present invention describes a purified or isolated nucleic acid
CC sequence (I) consisting essentially of one of the 93 nucleotide sequences
CC given in AAH81202 to AAH81294, where expression of the nucleic acid in a
CC microorganism is capable of inhibiting proliferation of a microorganism.
CC (I) have antibacterial and antibiotic activities, and can be used in gene
CC therapy. Expression of (I) in a microorganism inhibits proliferation of
CC the microorganism, and the manufactured antibiotic is useful for reducing
CC the activity or level of a gene product required for proliferation of a
CC microorganism in a subject, specifically humans. The nucleic acids that
CC inhibit bacterial growth or proliferation can be used as antisense
CC therapeutics for killing bacteria. In addition to therapeutic
CC applications, the nucleic acid sequences complementary to sequences
CC required for proliferation can be used as diagnostic tools. For example,
CC nucleic acid probes complementary to proliferation-required sequences
CC that are specific for particular species of microorganisms can be used as
CC probes to identify particular microorganism species in clinical
CC specimens. AAH81295 to AAH81487 encode the Escherichia coli proteins
CC given in AAG98239 to AAG98431, and AAH81488 to AAH81491 represent
CC oligonucleotides, which are used in the exemplification of the present
CC invention
XX Sequence 355 AA;
SQ

Query Match 95.9%; Score 1803; DB 4; Length 355;
Best Local Similarity 99.2%; Pred. No. 2.5e-171;
Matches 352; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 2 MSGGCKLWVWAGTGGHVPGLAVAHLLMAQGWQVRLGTADRMEDLVPKHGIEDFI 61
DB 1 MSGGCKLWVWAGTGGHVPGLAVAHLLMAQGWQVRLGTADRMEDLVPKHGIEDFI 60
QY 62 RISGLRGKIKALIAAPLRIFNWQRAIMKAYKPDVVLGMGGYVSGPGLAWSLIGIP 121
DB 61 RISGLRGKIKALIAAPLRIFNWQRAIMKAYKPDVVLGMGGYVSGPGLAWSLIGIP 120
QY 122 VVLEHQNGIAGLTNKLWARIATKQWQRAEAFNAEVVGNPVRTDVLALPLPOORLAGRE 181
DB 121 VVLEHQNGIAGLTNKLWARIATKQWQRAEAFNAEVVGNPVRTDVLALPLPOORLAGRE 180

QY 182 GPRVVLVVGSGQARILNQTMPQVAAKLGDSVLIWHQSGKSGSQSQVEQYAEAGQPQHKV 241
DB 181 GPRVVLVVGSGQARILNQTMPQVAAKLGDSVLIWHQSGKSGSQSQVEQYAEAGQPQHKV 240
QY 242 TEFIDDDMAAYANADVVCVSGALTVSEIAAAGLPALFVFPQHKDRQQYWNALPLEKAGA 301
DB 241 TEFIDDDMAAYANADVVCVSGALTVSEIAAAGLPALFVFPQHKDRQQYWNALPLEKAGA 300
QY 302 AKIIIEQQLSVDVANTLAGWSRETLTMAERARAASIPDATERVANEVSRVARA 356
DB 301 AKIIIEQQLSVDVANTLAGWSRETLTMAERARAASIPDATERVANEVSRVARA 355

RESULT 4

ABU28491
ID ABU28491 standard; protein; 355 AA.

XX AC ABU28491;

DT 19-JUN-2003 (first entry)

DE Protein encoded by Prokaryotic essential gene #14018.

KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

OS Escherichia coli.

XX PN W0200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;

XX DR N-PSDB; ACA32361.

XX PT WPI; 2003-029926/02.

XX PT New antisense nucleic acids, useful for identifying proteins or screening

XX PT for homologous nucleic acids required for cellular proliferation to

XX PT isolate candidate molecules for rational drug discovery programs.

XX PS Claim 25; SEQ ID NO 56415; 1766pp; English.

XX CC The invention relates to an isolated nucleic acid comprising any one of
XX CC the 5213 antisense sequences given in the specification where expression
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
XX CC encoding a polypeptide whose expression is inhibited by the antisense
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX CC polypeptide or its fragment whose expression is inhibited by the
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX CC proliferation or the activity of a gene in an operon required for
XX CC proliferation; (7) identifying a compound that influences the activity of
XX CC the gene product or that has an activity against a biological pathway
XX CC required for proliferation, or that inhibits cellular proliferation; (8)
XX CC identifying a gene required for cellular proliferation or the biological
XX CC pathway in which a proliferation-required gene or its gene product lies
XX CC or a gene on which the test compound that inhibits proliferation of an
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX CC compound's activity; (11) a culture comprising strains in which the gene
XX CC product is overexpressed or underexpressed; (12) determining the extent

CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 355 AA;

Query Match

Best Local Similarity 95.9%; Score 1803; DB 6; Length 355;

Matches 352; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 MSGQKRLMNVAGGTGGHVFPGLAVAHHLMAQGVQVWLGTDADRMADLVPKHGIEIDFI 61

DB 1 MSGQKRLMNVAGGTGGHVFPGLAVAHHLMAQGVQVWLGTDADRMADLVPKHGIEIDFI 60

QY 62 RISGLRGKIGIKALIAAPLIRFNARQARAIWKAYKPDVILGMGGYVSGPGLAAMSIGIP 121

DB 61 RISGLRGKIGIKALIAAPLIRFNARQARAIWKAYKPDVILGMGGYVSGPGLAAMSIGIP 120

QY 122 VTLHEQNGIAGLTNKLARIATKYNQABPGAFPAEYVVGNPVRTDYLALPLPQORLAGRE 181

DB 121 VTLHEQNGIAGLTNKLARIATKYNQABPGAFPAEYVVGNPVRTDYLALPLPQORLAGRE 180

QY 182 GPRVVLVVGSGQARILNQTMPQVAAKLGDSVLIWHQSGKSGSQSQVEQYAEAGQPQHKV 241

DB 181 GPRVVLVVGSGQARILNQTMPQVAAKLGDSVLIWHQSGKSGSQSQVEQYAEAGQPQHKV 240

QY 242 TEFIDDDMAAYANADVVCVSGALTVSEIAAAGLPALFVFPQHKDRQQYWNALPLEKAGA 301

DB 241 TEFIDDDMAAYANADVVCVSGALTVSEIAAAGLPALFVFPQHKDRQQYWNALPLEKAGA 300

QY 302 AKIIIEQQLSVDVANTLAGWSRETLTMAERARAASIPDATERVANEVSRVARA 356

DB 301 AKIIIEQQLSVDVANTLAGWSRETLTMAERARAASIPDATERVANEVSRVARA 355

RESULT 5

ABU47265

ID ABU47265 standard; protein; 355 AA.

XX AC ABU47265;

XX DT 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #32792.

XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX OS Salmomella typhimurium.

XX PN W0200277183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WO-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-MAR-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX MPI; 2003-029926/02.
DR N-FSDB; ACAS1135.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX
PS Claim 25; SEQ ID NO 75189; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 355 AA;

Query Match 89.7%; Score 1687; DB 6; Length 355;
Best Local Similarity 91.2%; Pred. No. 9.9e-160;
Matches 323; Conservative 18; Mismatches 13; Indels 0; Gaps 0;

QY 2 MSGGKELVMAGTGGHVPFGLAVAHHLMAQGWVRLGTADRMADLVPKHGIEDFI 61
DB 1 MSGQKRLVMAGTGGHVPFGLAVAHHLMAQGWVRLGTADRMADLVPKHGIEDFI 60
QY 62 RISLGRKGKIGKIALIAPLRIFNARQARAIMKAYKPDVVLGMGGYVSGPGLAANSLGIP 121
DB 61 RISLGRKGKIGKIALIAPLRIFNARQARAIMKAYKPDVVLGMGGYVSGPGLAANSLGIP 120
QY 122 VULHEQNGIAGLTKMLARIAIKWQAEFGAPFNAEIVGNPVRTDVLPLPQOBLAGE 181
DB 121 VULHEQNGIAGLTKMLARIAIKWQAEFGAPFNAEIVGNPVRTDVLPLPQOBLAGE 180
QY 182 GPRVRLVVGSGQARTLNQTPQVAAKLGDVLIWHQSGKGSQSVQEQAYAGAGPQHKV 241
DB 181 GPRVRLVVGSGQARTLNQTPQVAAKLGDVLIWHQSGKGSQSVQEQAYAGAGPQHKV 240
QY 242 TEFIDNMAAAYAWADVVCRCGALTIVSEIAAGLPAIFVFPQHKDRQQYWNALPLEKAGA 301
DB 241 TEFIDNMAAAYAWADVVCRCGALTIVSEIAAGLPAIFVFPQHKDRQQYWNALPLENAGA 300
QY 302 AKTIEQPOLSVDAVANTLAGWRETTLLTWAERARASIPDATERVANEVSRYAR 355
DB 301 AKTIEQPOLSVDAVANTLAGWRETTLLTWAERARASIPDATERVANEVSRYAR 354

RESULT 6

AAU38485
ID AAU38485 standard; protein; 355 AA.
XX
AC AAU38485;
XX
DT 14-FEB-2002 (first entry)
XX
DE Salmonella typhi cellular proliferation protein #376.
XX
KW Antisense; prokaryotic cellular proliferation protein; antibiotic;
KW antibacterial; drug design.
XX
OS Salmonella typhi.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US009180.
XX
PR 21-MAR-2000; 2000US-0191078P.
PR 23-MAY-2000; 2000US-0206848P.
PR 26-MAY-2000; 2000US-0207272P.
PR 23-OCT-2000; 2000US-0242578P.
PR 27-NOV-2000; 2000US-0253625P.
PR 22-DEC-2000; 2000US-0257931P.
PR 16-FEB-2001; 2001US-0269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
XX MPI; 2001-611495/70.
DR N-FSDB; AAS56344.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.
XX
PS Example 3; SEQ ID NO 14078; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes
CC themselves and the encoded proteins. The prokaryotes used are *Escherichia*
CC *coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*,
CC *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
CC antibodies capable of binding to the expressed proteins. The proteins can
CC be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous
CC nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence represents an essential prokaryotic
CC cellular proliferation protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 355 AA;

Query Match 89.2%; Score 1677; DB 4; Length 355;
Best Local Similarity 90.7%; Pred. No. 9.9e-159;
Matches 321; Conservative 19; Mismatches 14; Indels 0; Gaps 0;

QY 2 MSGGKELVMAGTGGHVPFGLAVAHHLMAQGWVRLGTADRMADLVPKHGIEDFI 61
DB 1 MSGQKRLVMAGTGGHVPFGLAVAHHLMAQGWVRLGTADRMADLVPKHGIEDFI 60
QY 62 RISLGRKGKIGKIALIAPLRIFNARQARAIMKAYKPDVVLGMGGYVSGPGLAANSLGIP 121

Db 61 RISLGRGKGVKALLAALPLRIFNAWQARAIMKRFDPDVLVGMGVSGPGLAANSLGIP 120
Qy 122 VVLHEQNGIAGLTNKLARIATKVMQAEPCAPNVEVGNVPTDVLALPQORLAGRE 181
Db 121 VVLHEQNGIAGLTNQLAKIATVMQAEPCAPNVEVGNVPTDVLALPQORLAGRD 180
Qy 182 GPRVLVVGSGQARILNQTMPQAAKLDGSDVLIHQSGKSGSQSQVQAYAEAGOPQHKV 241
Db 181 GPRVLVVGSGQARVLTNQTMPQAAKLDGSDVLIHQSGKSGQALTVQAYAGAGOPQHKV 240
Qy 242 TEFIDMAAAYAWADVVCVRSGLATVSEIAAAGLPALFVFFQHKDQOQYWNALPLEKAGA 301
Db 241 TEFIDMAAAYAWADVVCVRSGLATVSEIAAAGLPALFVFFQHKDQOQYWNALPLENAGA 300
Qy 302 AKIIEQPQLSDAVANTLAGSRETLTWAERARAASIPDATERVANEVSRRVAR 355
Db 301 AKIIEQPQFTVEAVADTLAGSREALLTWAERARAVSIPDATERVASEVSRRVAR 354

RESULT 7
ABU48097
ID ABU48097 standard; protein; 355 AA.
XX AC ABU48097;
XX 19-JUN-2003 (first entry)
XX Protein encoded by Prokaryotic essential gene #33624.
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX Salmomella typhi.
XX WO200277183-A2.
XX 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
XX 06-SEP-2001; 2001US-00948993.
XX 25-OCT-2001; 2001US-0342923P.
XX 08-FEB-2002; 2002US-00072851.
XX 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX N-PSDB; ACA51967.

XX New Antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX Claim 25; SEQ ID NO 76021; 17659p; English.
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the gene in an operon required for
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway
XX required for proliferation, or that inhibits cellular proliferation; (8)

CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 355 AA;

Query Match 89.2%; Score 1677; DB 6; Length 355;
Best Local Similarity 90.7%; Pred. No. 9.9e-159;
Matches 321; Conservative 19; Mismatches 14; Indels 0; Gaps 0;
Qy 2 MSGQKRLVMAGTGGHVPPGLAVAHHLMAQGWVRLGTADRMADLVPKHGIDIFI 61
Db 1 MSGQKRLVMAGTGGHVPPGLAVAHHLMAQGWVRLGTADRMADLVPKHGIDIFI 60
Qy 62 RISLGRGKGVKALLAALPLRIFNAWQARAIMKRFDPDVLVGMGVSGPGLAANSLGIP 121
Db 61 RISLGRGKGVKALLAALPLRIFNAWQARAIMKRFDPDVLVGMGVSGPGLAANSLGIP 120
Qy 122 VVLHEQNGIAGLTNKLARIATKVMQAEPCAPNVEVGNVPTDVLALPQORLAGRE 181
Db 121 VVLHEQNGIAGLTNQLAKIATVMQAEPCAPNVEVGNVPTDVLALPQORLAGRD 180
Qy 182 GPRVLVVGSGQARILNQTMPQAAKLDGSDVLIHQSGKSGSQSQVQAYAEAGOPQHKV 241
Db 181 GPRVLVVGSGQARVLTNQTMPQAAKLDGSDVLIHQSGKSGQALTVQAYAGAGOPQHKV 240
Qy 242 TEFIDMAAAYAWADVVCVRSGLATVSEIAAAGLPALFVFFQHKDQOQYWNALPLEKAGA 301
Db 241 TEFIDMAAAYAWADVVCVRSGLATVSEIAAAGLPALFVFFQHKDQOQYWNALPLENAGA 300
Qy 302 AKIIEQPQLSDAVANTLAGSRETLTWAERARAASIPDATERVANEVSRRVAR 355
Db 301 AKIIEQPQFTVEAVADTLAGSREALLTWAERARAVSIPDATERVASEVSRRVAR 354

RESULT 8
ABU31657
ID ABU31657 standard; protein; 348 AA.
XX AC ABU31657;
XX 19-JUN-2003 (first entry)
XX Protein encoded by Prokaryotic essential gene #17184.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX Klebsiella pneumoniae.
XX WO200277183-A2.
XX 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
XX 06-SEP-2001; 2001US-00948993.
XX 25-OCT-2001; 2001US-0342923P.

CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
CX
XX Sequence 348 AA;

Query Match 82.4%; Score 1549.5; DB 6; Length 348;
Best Local Similarity 86.3%; Pred. No. 5.5e-146;
Matches 302; Conservative 16; Mismatches 30; Indels 1; Gaps 1;
QY 2 MSGQKRLVMAGTGGHVPFGLAVAHLMACQWVRMLGTADRMEDLVPRKHGIEIDFI 61
Db 1 MSGQKRLVMAGTGGHVPFGLAVAHLMACQWVRMLGTADRMEDLVPRKHGIEIDFI 60
QY 62 RISGLRGKIKALIAAPLRIFNAWRQARIMKAYKPDVVLGMGGYVSGPGLAASLGIP 121
Db 61 RISGLRGKIKALIAAPLRIFNAWRQARIMKAYKPDVVLGMGGYVSGPGLAASLGIP 120
QY 122 VVLHEQNGIAGLTNKLARIATKVMQAEFGAPNAEVGNPVRTDVLALPLPQORLAGRE 181
Db 121 VVLHEQNGIAGLTNKLARIATKVMQAEFGAPNAEVGNPVRTDVLALPLPQORLAGRD 180
QY 182 GPRVRLVVGSGQARILNCTMPQVAALGDSVIIHQSGKGSQSQVQAYABAGOPQHKV 241
Db 181 GPRVRLVVGSGQARILNCTMPQVAALGDSVIIHQSGKGSQQLTVQAYAGTGPQHKV 240
QY 242 TEFDIDMAAYAWADVVCESGALTYSEIAAGLPALFPVFOHKDQOQYWNALPLEKAGA 301
Db 241 TEFDIDMAAYAWADVVCESGALTYSEIAAGLPALFPVFOHKDQOQYWNALPLENAGA 300
QY 302 AKIEQPQLSDVAVANTLAGWSRETLITMAERARAAASIPDATERVANEV 350
Db 301 AKIEQPQFTVDSVPT-GGVAAEALLTWAEVPAASVDPATERVASEV 348

RESULT 10
ABU50002
ID ABU50002 standard; protein; 356 AA.
XX
AC ABU50002;
XX
DT 19-JUN-2003 (first entry)
DE
DE Protein encoded by Prokaryotic essential gene #35529.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Yersinia pestis.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-0299926/02.
DR N-PSDB; ACA53872.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
PT

PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 77926; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX

Sequence 356 AA;

Query Match 78.3%; Score 1471; DB 6; Length 356;
Best Local Similarity 78.3%; Pred. No. 4e-138;
Matches 278; Conservative 36; Mismatches 41; Indels 0; Gaps 0;
QY 2 MSGQKRLVMAGTGGHVPFGLAVAHLMACQWVRMLGTADRMEDLVPRKHGIEIDFI 61
Db 1 MSGQKRLVMAGTGGHVPFGLAVAHLMACQWVRMLGTADRMEDLVPRKHGIEIDFI 60
QY 62 RISGLRGKIKALIAAPLRIFNAWRQARIMKAYKPDVVLGMGGYVSGPGLAASLGIP 121
Db 61 RISGLRGKIKALIAAPLRIFNAWRQARIMKAYKPDVVLGMGGYVSGPGLAASLGIP 120
QY 122 VVLHEQNGIAGLTNKLARIATKVMQAEFGAPNAEVGNPVRTDVLALPLPQORLAGRE 181
Db 121 VVLHEQNGIAGLTNKLARIATKVMQAEFGAPNAEVGNPVRTDVLALPLPQORLAGRE 180
QY 182 GPRVRLVVGSGQARILNCTMPQVAALGDSVIIHQSGKGSQSQVQAYABAGOPQHKV 241
Db 181 GPRVRLVVGSGQARILNCTMPQVAALGDSVIIHQSGKGSQQLTVQAYAGTGPQHKV 240
QY 242 TEFDIDMAAYAWADVVCESGALTYSEIAAGLPALFPVFOHKDQOQYWNALPLEKAGA 301
Db 241 TEFDIDMAAYAWADVVCESGALTYSEIAAGLPALFPVFOHKDQOQYWNALPLEKAGA 300
QY 302 AKIEQPQLSDVAVANTLAGWSRETLITMAERARAAASIPDATERVANEVSRVARA 356
Db 301 AKIEQPQFTATSVSLLASWDRATLLSMAERARSVAIPDATERVAEVAASKS 355

RESULT 11
ABM68867
ID ABM68867 standard; protein; 356 AA.
XX
XX ABM68867;
XX

20-NOV-2003 (first entry)
 Photorhabdus luminescens protein sequence #1964.
 Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
 detection; food; gene expression; plant; animal; microorganism; toxin;
 antibiotic; biopesticide; virulence factor; disease model; plague;
 whooping cough.
 Photorhabdus luminescens.
 WO200294867-A2.
 28-NOV-2002.
 07-FEB-2002; 2002WO-IB003040.
 07-FEB-2001; 2001FR-00001659.
 (INSP) INST PASTEUR.
 (CNRS) CNRS CENT NAT RECH SCI.
 Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
 Buchrieser C;
 WPI; 2003-148459/14.
 Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
 useful e.g. as therapeutic antimicrobials and agricultural pesticides.
 Claim 2; SEQ ID NO 1964; 1205pp; French.
 The invention relates to the isolation of genes and their encoded
 proteins from Photorhabdus luminescens. The isolated sequences are
 sources of probes and primers for detecting the genome of P. luminescens
 and related species; to study polymorphisms; for gene analysis and for
 detection/amplification of the genes. Antibodies (Ab) raised against the
 polypeptides encoded by the genes are used for detection/identification
 of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
 carry a gene-containing vector are used to select compounds that
 modulate, regulate, induce or inhibit expression of the genes in plants,
 animals or microorganisms other than P. luminescens and are able to alter
 response or sensitivity to toxins and antibiotics produced by P.
 luminescens. Cells transformed to express the genes are useful for
 recombinant production of the proteins, particularly toxins and
 antibacterials useful as insecticides, bactericides and fungicides. The
 genes, proteins, vectors containing the genes and Ab are also useful
 therapeutically (to treat microbial infection by bacteria or fungi that
 are sensitive to P. luminescens-encoded toxins or antibiotics) and as
 biopesticides. Other uses of the genes and the proteins are as virulence
 factors and for identifying targets of human diseases for which P.
 luminescens is a model (particularly plague and whooping cough). This
 sequence represents one of the isolated P. luminescens proteins

Query Match 76.1%; Score 1431; DB 6; Length 356;
 Best Local Similarity 76.3%; Pred. No. 4e-134;
 Matches 264; Conservative 44; Mismatches 38; Indels 0; Gaps 0;

QY 2 MSQGRKLMVAGTGGHVPFGLAVHLLMAQGVQVRLGTADRMEDLVPRKHGIEIDFI 61
 DB 1 MSGKTRLMVAGTGGHVPFGLAVHLLMAQGVQVRLGTADRMEDLVPRKHGIDIEFI 60
 QY 62 RISLRGKGKIALIAAPLRIFENAWROARIMKAYKEDVVLGMGYVSGGGLAANSLGIP 121
 DB 61 QISLRGKGKIALIAAPLRIFENAWROARIMKAYKEDVVLGMGYVSGGGLAANSLGIP 120
 QY 122 VVLHEQNGIAGLNKMLARIATKVMQAEPCAFNPAEVGNVTRDVLALPLPQORLAGRE 181
 DB 121 VVLHEQNGIAGLNKMLARIATKVMQAEPCAFNPAEVGNVTRDVLALPLPQORLAGRE 180
 QY 182 GPVRVLVVGSGQCARILNQAMPPEIARMGDKITLWHQTKGAKESVQYAYNSVACEHKI 240

DB 181 GPVRVLVVGSGQCARILNQAMPPEIARMGDKITLWHQTKGAKESVQYAYNSVACEHKI 240
 QY 242 TFFIDMAAAYAWADVVCVRSALTVSIAAAGLPALFVFPQHKDRQYWNALPLEKAGA 301
 DB 241 TFFIDMAQAWADWICVRSALTVSIAAAGLPALFVFPQHKDRQYWNALPLEKAGA 300
 QY 302 AKIIEQPOLSDVAVANTLWAGSEETLLTMAEPAASIPDATEVA 347
 DB 301 AKILEQPOFTVDVIELLTQWOPOLLEMAEKARSAAIVDATEQVS 346

RESULT 12

ABU40887
 ID ABU40887 standard; protein; 360 AA.
 XX
 AC ABU40887;
 DT 19-JUN-2003 (first entry)
 DE Protein encoded by Prokaryotic essential gene #26414.
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 OS Proteus sp.
 XX WO200277183-A2.
 PN 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX (ELIT-) ELITRA PHARM INC.
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX WPI; 2003-029926/02.
 DR N-PSDB; ACA44757.
 XX
 PS New antisense nucleic acids, useful for identifying proteins or screening
 for homologous nucleic acids required for cellular proliferation to
 isolate candidate molecules for rational drug discovery programs.
 Claim 25; SEQ ID NO 68811; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of
 the 6213 antisense sequences given in the specification where expression
 of the nucleic acid inhibits proliferation of a cell. Also included are:
 (1) a vector comprising a promoter operably linked to the nucleic acid
 encoding a polypeptide whose expression is inhibited by the antisense
 nucleic acid; (2) a host cell containing the vector; (3) an isolated
 polypeptide or its fragment whose expression is inhibited by the
 antisense nucleic acid; (4) an antibody capable of specifically binding
 the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 proliferation or the activity of a gene in an operon required for
 proliferation; (7) identifying a compound that influences the activity of
 the gene product or that has an activity against a biological pathway
 required for proliferation, or that inhibits cellular proliferation; (8)
 identifying a gene required for cellular proliferation or the biological
 pathway in which a proliferation-required gene or its gene product lies
 or a gene on which the test compound that inhibits proliferation of an
 organism acts; (9) manufacturing an antibiotic; (10) profiling a
 compound's activity; (11) a culture comprising strains in which the gene
 product is overexpressed or underexpressed; (12) determining the extent
 to which each of the strains is present in a culture or collection of
 strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 360 AA;

Query Match 71.5%; Score 1345.5; DB 6; Length 360;
Best Local Similarity 72.1%; Pred. No. 1.5e-125;
Matches 259; Conservative 40; Mismatches 55; Indels 5; Gaps 1;

QY 2 MSQGGKRLVMAGTGGHVPFGLAVAHILMAQGVRLGTADRMEDLVPKHGIEIDFI 61
Db 1 MSERKRLVMAGTGGHVPFGLAVAHILMAQGVRLGTADRMEDLVPKHGIEIYI 60
QY 62 RISLGRKGKIKALIAAPLRFNARQARIMKAYKPDVVLGMGGYVSGPGLAASLGIP 121
Db 61 RISLGRKGKIKALIAAPLRFNARQARIMKAYKPDVVLGMGGYVSGPGLAASLGIP 120
QY 122 VILHEONGIAGLTKNLARIATKVMQAEPCGAFNAPVWGNPVRTDVLALPLPOORLAGRE 181
Db 121 VILHEONGIAGLTKNLARIATKVMQAEPCGAFNAPVWGNPVRTDVLALPLPOORLAGRE 180
QY 182 GPRVRLVVGSGQARILNQTMPQVAALGDSVLIHQSGKGSQOSVEQAY-----AAGQ 236
Db 181 GAVRLVVGSGQARILNQTMPQVAALGDSVLIHQSGKGSQOSVEQAY-----AAGQ 240
QY 237 PQHKVTEFIDDMAAAVADVVYCRSGALTVSIIAAGLPALFVPPQHKDROQYWNALPL 296
Db 241 SEYKVTIEFIDIAQAYQADVVYCRSGALTVSIIAAGLPALFVPPQHKDROQYWNALPL 300
QY 297 EKAGAKIIEQPI.SYDVAANTLAGMSRETLITMAERARAASTPDATERVANEVSRAV 355
Db 301 ENAGAKIIEQNDITPEATITLNDWRHQLMLMAEKAGSAITDATERVANVITEVAK 359

RESULT 13
ABU27914
ID ABU27914 standard; protein; 281 AA.
XX
AC ABU27914;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #13441.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Enterobacter cloacae.
XX
PN W0200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX

DR MPI; 2003-029926/02.
XX N-PSDB; ACA31784.
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 55938; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 281 AA;

Query Match 70.3%; Score 1323; DB 6; Length 281;
Best Local Similarity 91.3%; Pred. No. 1.8e-123;
Matches 253; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

QY 5 QGKRLVMAGTGGHVPFGLAVAHILMAQGVRLGTADRMEDLVPKHGIEIDFI 64
Db 3 QPKRLVMAGTGGHVPFGLAVAHILMAQGVRLGTADRMEDLVPKHGIEIDFI 62
QY 65 GLRGKGIKALIAAPLRFNARQARIMKAYKPDVVLGMGGYVSGPGLAASLGIPVL 124
Db 63 GLRGKGIKALIAAPLRFNARQARIMKAYKPDVVLGMGGYVSGPGLAASLGIPVL 122
QY 125 HEONGIAGLTKNLARIATKVMQAEPCGAFNAPVWGNPVRTDVLALPLPOORLAGRGPV 184
Db 123 HEONGIAGLTKNLARIATKVMQAEPCGAFNAPVWGNPVRTDVLALPLPOORLAGRGPV 182
QY 185 RVLVVGSGQARILNQTMPQVAALGDSVLIHQSGKGSQOSVEQAYAEAGCPQHKVTEF 244
Db 183 RVLVVGSGQARILNQTMPQVAALGDSVLIHQSGKGSQOSVEQAYAEAGCPQHKVTEF 242
QY 245 IDDMAAAVADVVYCRSGALTVSEIAAAGLPALFVP 281
Db 243 IDDMAAAVADVVYCRSGALTVSEIAAAGLPALFVP 279

RESULT 14
ABU49532
ID ABU49532 standard; protein; 354 AA.
XX
XX ABU49532;
AC

CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 354 AA;

Query Match 59.7%; Score 1123; DB 6; Length 354;
Best Local Similarity 63.1%; Pred. No. 2.6e-103;
Matches 222; Conservative 46; Mismatches 76; Indels 8; Gaps 2;
QY 2 MSGQKELVMAGTGGHVPFGLVAHLMAGQVRLMGLTADRMEDLVPHKGIIDIFI 61
DB 1 MSEQKRLVMAGTGGHVPFGLVAHLMAGQVRLMGLTADRMEDLVPHKGIIDIFI 60
QY 62 RISLGRKGKIKALIAAPLRFNARQARAIMKAYKPDVVLGMGGYVSGGGLAANSLGIP 121
DB 61 QISLGRKGKIKALIAAPLRFNARQARAIMKAYKPDVVLGMGGYVSGGGLAANSLGIP 120
QY 122 VULHEONGIAGLTKMLARIATKVMQAEPCGAPFNAEVGNVPTDVLPLPQORLAGRE 181
DB 121 VILHEQNAVAGLNTSLAKIARRVLQAFPTAFPAEAVGNVPTDVLPLPQORLAGRE 180
QY 182 GPRVRLVWVGSGQARILNQTMPQVAAKLGDSVLIWHQSGKSGQSQVQAYAEAGQPH-- 239
DB 181 GXLRLVWVGSGQARILNQTLPNVAAQLSDKLEVRHVGQGAVEQVTTLY----PEHAS 235
QY 240 -KYTFEIDMAAYADVVVCRSGALTYSSEIAAGLPALFPFQHKDQOQYNALPLEK 298
DB 236 VTITFEIDMADAYADIVICRSGALTYSSEIAAGLPALFPFQHKDQOQYNALPLEK 295
QY 299 AGAAKIIQPOLSVDAVANTLWAGSRETLTMAERARAASIPDATERVANEV 350
DB 296 AGAATIIQPOLSVDAVANTLWAGSRETLTMAERARAASIPDATERVANEV 347

RESULT 16
AAU35596
ID AAU35596 standard; protein; 351 AA.
XX
AC AAU35596;
XX
DT 14-FEB-2002 (first entry)
XX
DE Haemophilus influenzae cellular proliferation protein #237.
XX
KW Antisense; prokaryotic cellular proliferation protein; antibiotic;
KW antibacterial; drug design.
XX
OS Haemophilus influenzae.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US003180.
PF
XX 21-MAR-2000; 2000US-0191078P.
XX 23-MAY-2000; 2000US-0206848P.
PR 26-MAY-2000; 2000US-0207272P.
PR 23-OCT-2000; 2000US-0242578P.
PR 27-NOV-2000; 2000US-0253625P.
PR 22-DEC-2000; 2000US-0257931P.
PR 16-FEB-2001; 2001US-0269308P.
XX (ELIT-) ELITRA PHARM INC.
PA
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;
XX Yamamoto RT, Xu RH;
XX WPI; 2001-611495/70.
DR N-PSDB; AAS53455.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.
XX
XX Example 3; SEQ ID NO 11189; 511pp; English.
PS
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes
CC themselves and the encoded proteins. The prokaryotes used are *Escherichia*
CC *coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*,
CC *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
CC antibodies capable of binding to the expressed proteins. The proteins can
CC be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous
CC nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence represents an essential prokaryotic
CC cellular proliferation protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 351 AA;

Query Match 56.6%; Score 1064.5; DB 4; Length 351;
Best Local Similarity 61.5%; Pred. No. 1.8e-97;
Matches 214; Conservative 44; Mismatches 89; Indels 1; Gaps 1;
QY 5 QGKELVMAGTGGHVPFGLVAHLMAGQVRLMGLTADRMEDLVPHKGIIDIFI 64
DB 2 KNKLLVMAGTGGHVPFGLVAHLMAGQVRLMGLTADRMEDLVPHKGIIDIFI 61
QY 65 GLRGKGIKALIAAPLRFNARQARAIMKAYKPDVVLGMGGYVSGGGLAANSLGIPVL 124
DB 62 GLRGKGIKALIAAPLRFNARQARAIMKAYKPDVVLGMGGYVSGGGLAANSLGIPVL 121
QY 125 HEONGIAGLTKMLARIATKVMQAEPCGAPFNAEVGNVPTDVLPLPQORLAGREPV 184
DB 122 HEQNAIAGLTKMLGIATCVLQAFPTAFPAEAVGNVPTDVLPLPQORLAGREPV 181
QY 185 RVLVWVGSGQARILNQTMPQVAAKLGDSVLIWHQSGKSGQSQVQAYAEAGQPHKYTEF 244
DB 182 RVLVWVGSGQARILNQTMPQVAAKLGDSVLIWHQSGKSGQSQVQAYAEAGQPHKYTEF 240
QY 245 IDDMAAYADVVVCRSGALTYSSEIAAGLPALFPFQHKDQOQYNALPLEKAGAKI 304
DB 241 IDDMAAYADVVVCRSGALTYSSEIAAGLPALFPFQHKDQOQYNALPLEKAGAKI 300
QY 305 IEQPOLSVDAVANTLWAGSRETLTMAERARAASIPDATERVANEVSR 352
DB 301 IEQADLTPEILVNYLKNLTRENTLQALMALKATNSMFPNAORVAEVIKQ 348

RESULT 17
ABU30498
ID ABU30498 standard; protein; 351 AA.
XX
XX
XX ABU30498;
XX 19-JUN-2003 (first entry)
XX
XX Protein encoded by prokaryotic essential gene #16025.
XX
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
XX Haemophilus influenzae.
XX OS
XX WO200277183-A2.
XX PN
XX
XX 03-OCT-2002.
XX PD
XX
XX 21-MAR-2002; 2002WO-US009107.
XX PF
XX
XX 21-MAR-2001; 2001US-00815242.
XX PR
XX 06-SEP-2001; 2001US-00948993.
XX PR
XX 25-OCT-2001; 2001US-0342923P.
XX PR
XX 08-FEB-2002; 2002US-00072851.
XX PR
XX 06-MAR-2002; 2002US-0362699P.
XX PR
XX
XX (ELIT-) ELITRA PHARM INC.
XX FA
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX PI
XX
XX WPI; 2003-029926/02.
XX DR
XX N-PSDB; ACA34368.
XX DR
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX PT
XX
XX Claim 25; SEQ ID NO 58422; 1766pp; English.
XX PS
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX CC the 6213 antisense sequences given in the specification where expression
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
XX CC encoding a polypeptide whose expression is inhibited by the antisense
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX CC polypeptide or its fragment whose expression is inhibited by the
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX CC proliferation or the activity of a gene in an operon required for
XX CC proliferation; (7) identifying a compound that influences the activity of
XX CC the gene product or that has an activity against a biological pathway
XX CC required for proliferation, or that inhibits cellular proliferation; (8)
XX CC identifying a gene required for cellular proliferation or the biological
XX CC pathway in which a proliferation-required gene or its gene product lies
XX CC or a gene on which the test compound that inhibits proliferation of an
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX CC compound's activity; (11) a culture comprising strains in which the gene
XX CC product is overexpressed or underexpressed; (12) determining the extent
XX CC to which each of the strains is present in a culture or collection of
XX CC strains; or (13) identifying the target of a compound that inhibits the
XX CC proliferation of an organism. The antisense nucleic acids are useful for
XX CC identifying proteins or screening for homologous nucleic acids required
XX CC for cellular proliferation to isolate candidate molecules for rational
XX CC drug discovery programs, or for screening homologous nucleic acids
XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
XX CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
XX CC the target prokaryotic essential genes. Note: the sequence data for this
XX CC patent did not form part of the printed specification, but was obtained
XX CC in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 351 AA;
Query Match 56.6%; Score 1064.5; DB 6; Length 351;
Best Local Similarity 61.5%; Pred. No. 1.8e-97;
Matches 214; Conservative 44; Mismatches 89; Indels 1; Gaps 1;
QY 5 OGKRLMVMAGGTGGHVFPGGLAVAHHLMAQQWQVWLGTADTADMEADLVPKHGIEIDFIRIS 64
DB 2 KKKLLVMAGGTGGHVFPGGLAVAHHLMAQQWQVWLGTADTADMEADLVPKHGIEIDFIRIS 61
QY 65 GLRGKGIKALIAAPLIRIFNAWRQARAINKAYKDPVILGNGGVSGPGGLAASLIGIPVL 124
DB 62 GLRGKGIKALLNAPFAIPRAVLQAKTIQEEKPDVILGNGGVSGPGGLAASLIGIPVL 121
QY 125 HEONGIAGLTNKLARIATKVMQAEPCGAFNAEVGVNVRTDLALPLPQORLAGREGPV 184
DB 122 HEONAIAGLTNKLKLIATCVLQAFPTAFHAEVGVNVPREDLFEMPNDIRFSDREKL 181
QY 185 RVLVVGSGQGARILNQTMPQVAAKLGDSVVIWHSGKGSQSQSVEQAYABACQPQHKVTEF 244
DB 182 RVLVVGSGQGARVLNHTLPKVAQLADKLEFRHVGKGAVEEVSQLYGE-NLEQVKITEF 240
QY 245 IDDMAAYAWADVVCRSGLTVSIIAAGLPALFVPFHQKDRQOYKVALPLEKAGAAKI 304
DB 241 IDDMAAYAWADVVCRSGLTVSIIAAGLPALFVPFHQKDRQOYKVALPLEKAGAAKI 300
QY 305 IEQPOLSVDAVANTILAGWSRETLITMAERARAASIPDATERVANEVSR 352
DB 301 IEQADLTPEILLVNYLKNLTRENLQMLKAKTMSMENAQAQVAEVIKQ 348

RESULT 18
AAU99817
ID AAU99817 standard; protein; 351 AA.
XX
XX AC AAU99817;
XX
XX 07-OCT-2002 (first entry)
XX
XX Haemophilus influenzae membrane associated UDP-glycosyltransferase.
XX
XX MURG; membrane associated UDP-glycosyltransferase; antibiotic;
XX antimicrobial; modulator of glycosyltransferase activity; drug design;
XX UDP-glycosyltransferase; directed drug design; random drug design;
XX grid-based drug design.
XX
XX Haemophilus influenzae.
XX Key Location/Qualifiers
XX Region 11..16 /label= G-loop_1
XX Region 99..106 /label= G-loop_2
XX Region 187..193 /label= G-loop_3
XX
XX WO200190301-A2.
XX
XX 29-NOV-2001.
XX
XX 09-APR-2001; 2001WO-US011500.
XX
XX 17-MAY-2000; 2000US-0204930P.
XX
XX (UYPR-) UNIV PRINCETON.
XX
XX Walker S, Ha S;
XX WPI; 2002-171402/22.
XX
XX Novel composition comprising crystalline form of Murg protein, a membrane
XX -associated UDP-glycosyltransferase involved in peptidoglycan
XX biosynthesis, for determining ability of chemical compound to bind Murg
XX

(1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published/bct/sequences](http://wipo.int/pub/published/bct/sequences)

```

SQ      sequence 367 / AA;
Query Match      46.3%; Score 871; DB 6; Length 367;
Best Local Similarity 50.7%; Pred. No. 4.4e-78;
Matches 183; Conservative 57; Mismatches 115; Indels 5; Gaps 3
QY      2 MSGGCKLMVMAGGTGGHVFPGLVAHHLMAQGMQVWVLTADRMADLVPKHGIEIDFI 61
Db      1 MTSTQRTLMVMAGGTGGHVFPGLVAHRLMAQGMVWVVLGNPAGMEATLVPRHGIPMEYV 60
QY      62 RISGLRKGKIKALIAPLRIFNWRQBARIMKAYKPDVVLGHMGYVSGGGLAANSLGIP 121
Db      61 RFGLRGKGLATKALPFLNLLRACASLRALRRVRKPDVVLGHMGYITTFPAGLVLTWTCRP 120
QY      122 VILHQNGIAGLTNKLWARTATKVMQAEPCGAFNPAEWGNPRTDVIALPLPQRLAGRE 181
Db      121 LVLEQNSIAGLTNKLWAKLAKRVLVAFPCALPNAEWGNTPIRTEATPPQARYAARS 180

```

QY 182 GPVRLVVGSGGARILNQTMPQVAKLG--DSVLIWHQSGKSGSQSQVEQYAEAG---Q 236
Db 181 GKRLLVVGSGGAAALNEVPRALALAPDRPQVHQGAKHIDTLKENYEAAGLSCG 240
QY 237 PQHKVTEFIDDMAAYAWADVVCVRSGLTVSEIAAAGLPALFVFPQHK-DROQYWNALP 295
Db 241 SDVALVPFIDDMASAYANADLVICSGAMTVAEIAAVGVAALFVFPFHAVDHQTNAEF 300
QY 296 LEKAGAKIIEQPQLSVDAVANTLAGSRETLLTWAERARASTPDATERVANEVSAR 355
Db 301 LAEQGAALVQOQDLSAELADMLRQSGRDSLAWAERSRSLAKPDATDEVARCAVAG 360
QY 356 A 356
Db 361 A 361

RESULT 20

ABU21174
ID ABU21174 standard; protein; 372 AA.

XX AC ABU21174;
XX 19-JUN-2003 (first entry)

XX DE Protein encoded by Prokaryotic essential gene #6701.
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX OS Burkholderia fungorum.

XX PN W020027183-A2.

XX PD 03-OCT-2002.

XX PF 21-MAR-2002; 2002WC-US009107.

XX PR 21-MAR-2001; 2001US-00815242.

XX PR 06-SEP-2001; 2001US-00948993.

XX PR 25-OCT-2001; 2001US-0342923P.

XX PR 08-FEB-2002; 2002US-00072851.

XX PR 06-FEB-2002; 2002US-0362699P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX DR WPI; 2003-029926/02.

XX DR N-PSDB; ACA25044.

XX PS Claim 25; SEQ ID NO 49098; 1765pp; English.
XX CC The invention relates to an isolated nucleic acid comprising any one of
the 6213 antisense sequences given in the specification where expression
of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
encoding a polypeptide whose expression is inhibited by the antisense
nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC the gene product; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies

CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 372 AA;

QY Query Match 45.1%; Score 849; DB 6; Length 372;

Db Best Local Similarity 49.9%; Pred. No. 7.2e-76;

XX Matches 178; Conservative 60; Mismatches 113; Indels 6; Gaps 3;

QY 7 KELVVMAGGTGGHVFPGLVAHAHLMAQGWVRLGTADRMEDLVPKHGTEIDFIRISGL 66

Db 7 RFLVVMAGGTGGHVFPGLVAHAHLMAQGWVRLGTADRMEDLVPKHGTEIDFIRISGL 66

QY 67 RKGKIKALIAAPLIRIFNAWQARAIMKAYDPDVLGMGGYVSGPGLAANSLGIPVVLHE 126

Db 67 RKGKIKALIAAPLIRIFNAWQARAIMKAYDPDVLGMGGYVSGPGLAANSLGIPVVLHE 126

QY 127 QNGTAGLTKWLARIATKVMQAEPCAFENAVVGNPVRTDVLALPLPOQRLAGREGPVRV 186

Db 127 QNSIAGLTKWLARIATKVMQAEPCAFENAVVGNPVRTDVLALPLPOQRLAGREGPVRV 186

QY 187 LVVGSQSGARILNQTMPQVAKL--GDSVIIHQSGKSGSQSQVEQYAEAG---QOQKV 241

Db 187 LVVGSQSGARILNQTMPQVAKL--GDSVIIHQSGKSGSQSQVEQYAEAG---QOQKV 241

QY 242 TEFIDDMAAYAWADVVCVRSGLTVSEIAAAGLPALFVFPQHK-DROQYWNALPLEKAG 300

Db 242 TEFIDDMAAYAWADVVCVRSGLTVSEIAAAGLPALFVFPQHK-DROQYWNALPLEKAG 300

QY 247 VPFIDDMTSAYANADLVICSGAMTVSEISAVGVAALFVFPFHAVDHQTNAFLADNG 306

Db 247 VPFIDDMTSAYANADLVICSGAMTVSEISAVGVAALFVFPFHAVDHQTNAFLADNG 306

QY 301 AAKIIEQPQLSVDAVANTLAGSRETLLTWAERARASTPDATERVANEVSARAL 357

Db 301 AAKIIEQPQLSVDAVANTLAGSRETLLTWAERARASTPDATERVANEVSARAL 357

QY 307 AALVYQQRDLASAEIADLWLRSTQRTETLAEMASRSRLAKPDATDEQVAQICATVAGSI 363

Db 307 AALVYQQRDLASAEIADLWLRSTQRTETLAEMASRSRLAKPDATDEQVAQICATVAGSI 363

Search completed: June 7, 2004, 07:13:16

Job time : 65 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 07:12:05 ; Search time 22 seconds
(without alignments)
854.175 Million cell updates/sec

Title: US-09-829-275-1

Perfect score: 1891

Sequence: 1 MMSGGKRLMAGTGGHV.....RVANESVRVALEHHHHH 364

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Issued Patents RA:*

1: /cgn2_6/ptodata/2/iaa/5A.COMB.pap:*

2: /cgn2_6/ptodata/2/iaa/5B.COMB.pap:*

3: /cgn2_6/ptodata/2/iaa/5A.COMB.pap:*

4: /cgn2_6/ptodata/2/iaa/5B.COMB.pap:*

5: /cgn2_6/ptodata/2/iaa/PCTUS.COMB.pap:*

6: /cgn2_6/ptodata/2/iaa/backfiles1.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|-------|----------------------|
| 1 | 1803 | 95.9 | 355 | 3 | US-08-984-618-14 |
| 2 | 1629 | 86.6 | 370 | 4 | US-09-489-039A-12435 |
| 3 | 1345.5 | 71.5 | 360 | 4 | US-09-543-681A-7067 |
| 4 | 1064.5 | 56.6 | 351 | 3 | US-08-984-618-15 |
| 5 | 715.5 | 38.0 | 390 | 4 | US-09-252-991A-24124 |
| 6 | 701.5 | 37.3 | 397 | 4 | US-09-543-236-2678 |
| 7 | 680.5 | 36.2 | 366 | 4 | US-09-328-352-5036 |
| 8 | 422 | 22.4 | 363 | 3 | US-08-984-618-13 |
| 9 | 420 | 22.3 | 363 | 4 | US-09-107-532A-6757 |
| 10 | 411.5 | 21.9 | 370 | 4 | US-09-134-000C-4145 |
| 11 | 406.5 | 21.6 | 363 | 3 | US-08-984-618-16 |
| 12 | 284.5 | 15.1 | 357 | 4 | US-09-198-452A-977 |
| 13 | 273.5 | 14.5 | 352 | 2 | US-08-751-474-2 |
| 14 | 261 | 13.9 | 366 | 4 | US-09-134-001C-3198 |
| 15 | 131 | 7.0 | 525 | 3 | US-08-984-618-17 |
| 16 | 127 | 6.8 | 382 | 3 | US-08-984-618-12 |
| 17 | 122.5 | 6.5 | 572 | 4 | US-09-252-991A-26203 |
| 18 | 117 | 6.2 | 463 | 4 | US-08-311-731A-7 |
| 19 | 115.5 | 6.1 | 390 | 3 | US-09-036-987A-8 |
| 20 | 115.5 | 6.1 | 390 | 4 | US-09-370-700-8 |
| 21 | 115.5 | 6.1 | 390 | 4 | US-09-603-207-8 |
| 22 | 114.5 | 6.1 | 537 | 4 | US-09-653-270A-11 |
| 23 | 114.5 | 6.1 | 537 | 4 | US-09-651-941-11 |
| 24 | 114.5 | 6.1 | 537 | 4 | US-09-955-597-11 |
| 25 | 108.5 | 5.8 | 372 | 1 | US-08-196-218-33 |
| 26 | 108.5 | 5.8 | 372 | 1 | US-08-681-953-33 |
| 27 | 108.5 | 5.8 | 559 | 4 | US-08-506-296B-66 |

| | | | | | | |
|-----|-------|-----|-------|---|----------------------|--------------------|
| 28 | 108 | 5.7 | 743 | 4 | US-09-252-991A-26512 | Sequence 26512, A |
| 29 | 106 | 5.6 | 452 | 4 | US-08-311-731A-6 | Sequence 6, Appli |
| 30 | 106 | 5.6 | 1271 | 1 | US-08-095-734-2 | Sequence 2, Appli |
| 31 | 106 | 5.6 | 1271 | 1 | US-08-444-623-2 | Sequence 2, Appli |
| 32 | 106 | 5.6 | 1271 | 3 | US-08-471-863-2 | Sequence 2, Appli |
| 33 | 106 | 5.6 | 1271 | 4 | US-09-342-563-2 | Sequence 2, Appli |
| 34 | 106 | 5.6 | 1271 | 5 | PCT-US94-08267-2 | Sequence 2, Appli |
| 35 | 104.5 | 5.6 | 738 | 4 | US-09-252-991A-31185 | Sequence 31185, A |
| 36 | 103 | 5.5 | 746 | 4 | US-09-252-991A-18587 | Sequence 18587, A |
| 37 | 103 | 5.5 | 749 | 4 | US-09-252-991A-16977 | Sequence 16977, A |
| 38 | 102 | 5.4 | 1650 | 4 | US-09-252-991A-21798 | Sequence 21798, A |
| 39 | 101.5 | 5.4 | 945 | 4 | US-09-252-991A-30699 | Sequence 30699, A |
| 40 | 101.5 | 5.4 | 4545 | 2 | US-08-804-227C-14 | Sequence 14, Appli |
| 41 | 101.5 | 5.4 | 4550 | 2 | US-08-804-227C-18 | Sequence 8, Appli |
| 42 | 101.5 | 5.4 | 4550 | 2 | US-08-804-198-2 | Sequence 2, Appli |
| 43 | 101 | 5.4 | 496 | 4 | US-09-252-991A-31639 | Sequence 31639, A |
| 44 | 100.5 | 5.3 | 625 | 4 | US-09-252-991A-23005 | Sequence 23005, A |
| 45 | 100.5 | 5.3 | 3491 | 2 | US-07-642-734C-2 | Sequence 2, Appli |
| 46 | 100.5 | 5.3 | 3491 | 3 | US-08-439-009A-2 | Sequence 2, Appli |
| 47 | 100 | 5.3 | 288 | 4 | US-09-252-991A-23714 | Sequence 23714, A |
| 48 | 100 | 5.3 | 297 | 4 | US-08-755-307A-3 | Sequence 3, Appli |
| 49 | 99.5 | 5.3 | 488 | 4 | US-09-489-039A-13218 | Sequence 13218, A |
| 50 | 99.5 | 5.3 | 3170 | 3 | US-09-036-987A-4 | Sequence 4, Appli |
| 51 | 99.5 | 5.3 | 3170 | 3 | US-09-370-700-4 | Sequence 4, Appli |
| 52 | 99.5 | 5.3 | 3170 | 4 | US-09-603-207-4 | Sequence 4, Appli |
| 53 | 99 | 5.3 | 443 | 4 | US-09-252-991A-31911 | Sequence 31911, A |
| 54 | 98.5 | 5.2 | 379 | 4 | US-09-252-991A-21117 | Sequence 21117, A |
| 55 | 98.5 | 5.2 | 986 | 4 | US-08-311-731A-2 | Sequence 2, Appli |
| 56 | 98 | 5.2 | 381 | 4 | US-09-489-039A-10807 | Sequence 10807, A |
| 57 | 98 | 5.2 | 513 | 4 | US-09-252-991A-30773 | Sequence 30773, A |
| 58 | 98 | 5.2 | 1346 | 3 | US-08-804-227C-11 | Sequence 11, Appli |
| 59 | 98 | 5.2 | 1580 | 2 | US-08-804-198-5 | Sequence 5, Appli |
| 60 | 98 | 5.2 | 11877 | 3 | US-09-105-537-6 | Sequence 6, Appli |
| 61 | 98 | 5.2 | 348 | 4 | US-09-489-039A-10294 | Sequence 10294, A |
| 62 | 97.5 | 5.2 | 738 | 4 | US-09-252-991A-30721 | Sequence 30721, A |
| 63 | 97.5 | 5.2 | 828 | 4 | US-09-489-039A-12219 | Sequence 12219, A |
| 64 | 97 | 5.2 | 1151 | 4 | US-09-252-991A-21328 | Sequence 21328, A |
| 65 | 96 | 5.1 | 550 | 4 | US-09-252-991A-18287 | Sequence 18287, A |
| 66 | 96 | 5.1 | 679 | 4 | US-09-252-991A-18681 | Sequence 18681, A |
| 67 | 96 | 5.1 | 3729 | 2 | US-08-804-227C-4 | Sequence 4, Appli |
| 68 | 96 | 5.1 | 305 | 4 | US-09-489-039A-9594 | Sequence 9594, Ap |
| 69 | 95.5 | 5.1 | 417 | 4 | US-09-252-991A-28049 | Sequence 28049, A |
| 70 | 95.5 | 5.1 | 563 | 4 | US-09-252-991A-31048 | Sequence 31048, A |
| 71 | 95.5 | 5.1 | 852 | 4 | US-09-433-043B-118 | Sequence 118, App |
| 72 | 95.5 | 5.1 | 272 | 4 | US-09-252-991A-26742 | Sequence 26742, A |
| 73 | 95 | 5.1 | 349 | 1 | US-09-252-991A-31096 | Sequence 31096, A |
| 74 | 95 | 5.1 | 399 | 1 | US-08-096-623A-20 | Sequence 20, Appli |
| 75 | 95 | 5.1 | 455 | 4 | US-09-252-991A-17642 | Sequence 17642, A |
| 76 | 95 | 5.1 | 595 | 4 | US-09-252-991A-30780 | Sequence 30780, A |
| 77 | 95 | 5.1 | 735 | 4 | US-09-252-991A-30569 | Sequence 30569, A |
| 78 | 95 | 5.1 | 971 | 3 | US-09-107-149-19 | Sequence 19, Appli |
| 79 | 95 | 5.1 | 1074 | 2 | US-08-768-147B-2 | Sequence 2, Appli |
| 80 | 95 | 5.1 | 1074 | 3 | US-08-968-752B-2 | Sequence 2, Appli |
| 81 | 95 | 5.1 | 1074 | 3 | US-09-107-149-3 | Sequence 3, Appli |
| 82 | 95 | 5.1 | 1074 | 3 | US-09-536-224-2 | Sequence 2, Appli |
| 83 | 95 | 5.1 | 4928 | 3 | US-09-036-987A-5 | Sequence 5, Appli |
| 84 | 95 | 5.1 | 4928 | 3 | US-09-370-700-5 | Sequence 5, Appli |
| 85 | 95 | 5.1 | 4928 | 3 | US-09-603-207-5 | Sequence 5, Appli |
| 86 | 95 | 5.1 | 391 | 2 | US-08-759-561B-13 | Sequence 13, Appli |
| 87 | 94.5 | 5.0 | 391 | 2 | US-09-304-711-13 | Sequence 13, Appli |
| 88 | 94.5 | 5.0 | 391 | 4 | US-09-173-281-13 | Sequence 13, Appli |
| 89 | 94.5 | 5.0 | 540 | 4 | US-09-252-991A-19797 | Sequence 19797, A |
| 90 | 94.5 | 5.0 | 2035 | 1 | US-08-046-585-5 | Sequence 5, Appli |
| 91 | 94.5 | 5.0 | 2035 | 1 | PCT-US93-11721-5 | Sequence 5, Appli |
| 92 | 94.5 | 5.0 | 2035 | 1 | US-08-393-703-5 | Sequence 5, Appli |
| 93 | 94.5 | 5.0 | 3739 | 3 | US-09-320-678-2 | Sequence 2, Appli |
| 94 | 94.5 | 5.0 | 3739 | 3 | US-09-105-537-33 | Sequence 33, Appli |
| 95 | 94.5 | 5.0 | 3739 | 4 | US-09-141-908-3 | Sequence 3, Appli |
| 96 | 94.5 | 5.0 | 3739 | 4 | US-09-657-440-2 | Sequence 2, Appli |
| 97 | 94.5 | 5.0 | 280 | 4 | US-09-328-352-7042 | Sequence 7042, Ap |
| 98 | 94 | 5.0 | 711 | 1 | US-08-235-838-7 | Sequence 7, Appli |
| 99 | 94 | 5.0 | 711 | 2 | US-08-465-473B-7 | Sequence 7, Appli |
| 100 | 94 | 5.0 | 711 | 2 | | |

ALIGNMENTS

RESULT 1

US-08-984-618-14
; Sequence 14, Application US/08984618
; Patent No. 6251647
; GENERAL INFORMATION:
; APPLICANT: de Lencastre, Herminia
; APPLICANT: Tomasz, Alexander
; TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF
; TITLE OF INVENTION: METHICILLIN RESISTANT BACTERIA AND ANTAGONISTS THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/984,618
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-213 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: E. coli
; US-08-984-618-14

Query Match 95.9%; Score 1803; DB 3; Length 355;
Best Local Similarity 99.2%; Pred. No. 7.2e-179;
Matches 352; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 2 MSGGKRLMWAGTGGHVPFGLAVAHLLMAQGWVRLGTADRMEDLVPKHGIEIDFI 61
DB 1 MSGGKRLMWAGTGGHVPFGLAVAHLLMAQGWVRLGTADRMEDLVPKHGIEIDFI 60
QY 62 RISGLRGKIKALIAAPLRFNARQARATWKAYKDPDVLGMGGYVSGPGLAASLSGIP 121
DB 61 RISGLRGKIKALIAAPLRFNARQARATWKAYKDPDVLGMGGYVSGPGLAASLSGIP 120
QY 122 VVLEHQNGIAGLTNKLARIATKVMQAPGAFNPAEYVGNPVRTDVLALPLPQORLAGRE 181
DB 121 VVLEHQNGIAGLTNKLARIATKVMQAPGAFNPAEYVGNPVRTDVLALPLPQORLAGRE 180
QY 182 GPVRVLVVGSGQARIINQTMPOVAAKLGDSVTIWHOSGKGSGQSQSVQEAQVAGQPOHKV 241
DB 181 GPVRVLVVGSGQARIINQTMPOVAAKLGDSVTIWHOSGKGSGQSQSVQEAQVAGQPOHKV 240
QY 242 TEFTDDMAAAAYAWADVVCVRSGLTVSEIAAGLPALFVFPFHQKDRQQYWNALPLEKAGA 301

DB 241 TEFTDDMAAAAYAWADVVCVRSGLTVSEIAAGLPALFVFPFHQKDRQQYWNALPLEKAGA 300
QY 302 AKIIEQPOLSVDAVANTLAGWSRETLTMAERAPAAASIPDATERVANEVSRVARA 356
DB 301 AKIIEQPOLSVDAVANTLAGWSRETLTMAERAPAAASIPDATERVANEVSRVARA 355

RESULT 2

US-09-489-039A-12435
; Sequence 12435, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12435
; LENGTH: 370
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
; US-09-489-039A-12435

Query Match 86.6%; Score 1629; DB 4; Length 370;

Best Local Similarity 88.5%; Pred. No. 9.2e-161;
Matches 314; Conservative 19; Mismatches 22; Indels 0; Gaps 0;
QY 2 MSGGKRLMWAGTGGHVPFGLAVAHLLMAQGWVRLGTADRMEDLVPKHGIEIDFI 61
DB 15 MSGGKRLMWAGTGGHVPFGLAVAHLLMDQGWQVRLGTADRMEDLVPKHGIEIDFI 74
QY 62 RISGLRGKIKALIAAPLRFNARQARATWKAYKDPDVLGMGGYVSGPGLAASLSGIP 121
DB 75 LISGLRGKIKALIAAPLRFNARQARATWKAYKDPDVLGMGGYVSGPGLAASLSGIP 134
QY 122 VVLEHQNGIAGLTNKLARIATKVMQAPGAFNPAEYVGNPVRTDVLALPLPQORLAGRE 181
DB 135 VVLEHQNGIAGLTNKLARIATKVMQAPGAFNPAEYVGNPVRTDVLALPLPQORLAGRE 194
QY 182 GPVRVLVVGSGQARIINQTMPOVAAKLGDSVTIWHOSGKGSGQSQSVQEAQVAGQPOHKV 241
DB 195 GPVRVLVVGSGQARIINQTMPOVAAKLGDSVTIWHOSGKGSGQSQSVQEAQVAGQPOHKV 254
QY 242 TEFTDDMAAAAYAWADVVCVRSGLTVSEIAAGLPALFVFPFHQKDRQQYWNALPLEKAGA 301
DB 255 TEFTDDMAAAAYAWADVVCVRSGLTVSEIAAGLPALFVFPFHQKDRQQYWNALPLEKAGA 314
QY 302 AKIIEQPOLSVDAVANTLAGWSRETLTMAERAPAAASIPDATERVANEVSRVARA 356
DB 315 AKIIEQPOLSVDAVANTLAGWSRETLTMAERAPAAASIPDATERVANEVSRVARA 369

RESULT 3

US-09-543-681A-7067
; Sequence 7067, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7067
; LENGTH: 360
; TYPE: PRT

US-09-894-618-15
; ORGANISM: Proteus mirabilis
; ORIGINAL SOURCE: US-08-984-618-15
; HYPOTHETICAL: NO
; ORGANISM: Haemophilus influenzae
Query Match 71.5%; Score 1345.5; DB 4; Length 360;
Best Local Similarity 72.1%; Pred. No. 2.5e-131;
Matches 259; Conservative 40; Mismatches 55; Indels 5; Gaps 1;
QY 2 MSQGGKLMVAGTGGHVPGLAVAHLMQAQGVRLGTADRMADLVPKHGIEIDFI 61
DB 1 MSERKRLVMAGTGGHVPGLAVAHLMQAQGVRLGTADRMADLVPKHGIEIEI 60
QY 62 RISGLRGKIKALIAAPLRIFFNAWRQARAIKAYKDPVVLGMGGYVSGPGLAANSLGIP 121
DB 61 RISGLRGKIKALIAAPLRIFFNAWRQARAIKAYKDPVVLGMGGYVSGPGLAANSLGIP 120
QY 122 VILHEONGIAGLTKMLARIATKVMQAEPCGAFPAEAVVGNPVRTDVLALPLPQORLAGRE 181
DB 121 VILHEONGIAGLTKMLARIATKVMQAEPCGAFPAEAVVGNPVRTDVLALPLPQORLAGRE 180
QY 182 GPVRLVVGSGQARILNOTMPOVAAGLGDVLIWHQSGKGSQSVQAEV 236
DB 181 GAVRLVVGSGQARILNOTMPOVAAGLGDVLIWHQSGKGSQSVQAEV 240
QY 237 POHKVTEFIDDDAAAYAWADVVCRCGALTVEIAAGLPAALFVPPQHKDRQOYWNALPL 296
DB 241 SEYKVFIDDDIAQAYQWADVVCRCGALTVEIAAGLPAALFVPPQHKDRQOYWNALPL 300
QY 297 EKAGAAKIIQPOLSVDAVANTLAGRSRETLTMAERARAASIPDATERVANEVSVAR 355
DB 301 ENAGAARIIEQDLTPPAIDTLENWDRHQLMLAKQAQSVAITDATERVANVIEVAK 359

RESULT 4
US-08-984-618-15
; Sequence 15, Application US/08984618
; Patent No. 6251647
; GENERAL INFORMATION:
; APPLICANT: de Lencastre, Herminia
; APPLICANT: Tomasz, Alexander
; TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF
; TITLE OF INVENTION: METHICILLIN RESISTANT BACTERIA AND ANTAGONISTS THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; APPLICATION NUMBER: US/08/984,618
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCES/DOCKET NUMBER: 600-1-213 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 351 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-09-894-618-15
; HYPOTHETICAL: NO
; ORIGINAL SOURCE: US-08-984-618-15
; ORGANISM: Haemophilus influenzae
Query Match 56.6%; Score 1064.5; DB 3; Length 351;
Best Local Similarity 61.5%; Pred. No. 3.8e-102;
Matches 214; Conservative 44; Mismatches 89; Indels 1; Gaps 1;
QY 5 QCKRLVMAGTGGHVPGLAVAHLMQAQGVRLGTADRMADLVPKHGIEIDFIRIS 64
DB 2 KKKLLVMAGTGGHVPGLAVAHLMQAQGVRLGTADRMADLVPKHGIEIDFIRIS 61
QY 65 GLRGKIKALIAAPLRIFFNAWRQARAIKAYKDPVVLGMGGYVSGPGLAANSLGIPVL 124
DB 62 GLRGKIKALIAAPLRIFFNAWRQARAIKAYKDPVVLGMGGYVSGPGLAANSLGIPVL 121
QY 125 HEONGIAGLTKMLARIATKVMQAEPCGAFPAEAVVGNPVRTDVLALPLPQORLAGREGPV 184
DB 122 HEONGIAGLTKMLARIATKVMQAEPCGAFPAEAVVGNPVRTDVLALPLPQORLAGREGPV 181
QY 185 RVLVVGSGQARILNOTMPOVAAGLGDVLIWHQSGKGSQSVQAEVAGQPOHKVTEF 244
DB 182 RVLVVGSGQARILNOTMPOVAAGLGDVLIWHQSGKGSQSVQAEVAGQPOHKVTEF 240
QY 245 IDDMAAAYAWADVVCRCGALTVEIAAGLPAALFVPPQHKDRQOYWNALPLEKAGAAKI 304
DB 241 IDDMAAAYAWADVVCRCGALTVEIAAGLPAALFVPPQHKDRQOYWNALPLEKAGAAKI 300
QY 305 IEQPOLSVDAVANTLAGRSRETLTMAERARAASIPDATERVANEVSR 352
DB 301 IEQADLTPEILVNLKNTRENLLQALKAKTMSMPNAAQVAVIKQ 348

RESULT 5
US-09-252-991A-24124
; Sequence 24124, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24124
; LENGTH: 390
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24124
Query Match 38.0%; Score 715.5; DB 4; Length 390;
Best Local Similarity 44.2%; Pred. No. 8.2e-66;
Matches 157; Conservative 64; Mismatches 123; Indels 11; Gaps 6;
QY 9 LMVAGTGGHVPGLAVAHLMQAQGVRLGTADRMADLVPKHGIEIDFIRISGLRG 68
DB 38 VLMAGTGGHVPGLAVAHLMQAQGVRLGTADRMADLVPKHGIEIDFIRISGLRG 97
QY 69 KGKIALIAAPLRIFFNAWRQARAIKAYKDPVVLGMGGYVSGPGLAANSLGIPVLHEQN 128
DB 98 KGKIALIAAPLRIFFNAWRQARAIKAYKDPVVLGMGGYVSGPGLAANSLGIPVLHEQN 157
QY 129 GIAGLTKMLARIATKVMQAEPCGAFPAEAVVGNPVRTDVLALPLPQORLAGREGPV 185
DB 158 AVAGTANRSLAPIARVCEAFPDTPFASDKRLTGTGVPVAGELFLDAHARAPLTGR -VN 215
QY 186 VLVVGSGQARILNOTMPOVAAGLGDVLIWHQSGKGSQSVQAEVAGQPOHKVTE 243

Db 216 LVLGSLGAEPLNKLLEALAQVPLEIRPAIRHOAGQHAETIAERTVTA-VEADVAP 274
Qy 244 FIDDMAAYAWADVVCRSALTYSIAAAGLPALFPVPOHK-DROQYVNALPLEKAGAA 302
Db 275 FIDMATATWADLVICRAGALTIVSELTAAGLPALFVLPLPHADDDHOTENAEFLVRSAG 334
Qy 303 KITEQPLSDVAVANTLAG--WSRETLTMAERARAASIPDATERVANEVSRVAR 355
Db 335 RLLPQKSTGAELAQLSEVLMHPETLRSWADQARSLEAKPEATRITVWDACLEVAR 369

RESULT 6
US-09-540-236-2678
; Sequence 2678, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2678
; LENGTH: 397
; TYPE: PRT
; ORGANISM: M.catarhalis
US-09-540-236-2678

Query Match 37.3%; Score 701.5; DB 4; Length 397;
Best Local Similarity 43.9%; Pred. No. 2.4e-64;
Matches 157; Conservative 65; Mismatches 121; Indels 15; Gaps 6;

Qy 9 LVMWAGTGGHVPFGLAVAHHLMAQGVWLTADRMADLPVKGIGIEDIRISGLRG 68
Db 37 VLWMAAGTGGHVPFGLAVAHHLMAQGVWLTADRMADLPVKGIGIEDIRISGLRG 96
Qy 69 KGKICALIAPLRIFNARQARIMKAYKPDVVLGMGGYVSGPGGLAAWSLGLIPVVLHEQ 128
Db 97 KGLVRAIKLPFLMFKVAMSKNIKKNNIDVIGFGYVTFAPGLAKLCKPLIIEHQ 156
Qy 129 GIAGLTNKLARIATKYMQAEPGAPFN-----AEVGNPVRTDVLALPLPQORLAGREGV 184
Db 157 AIAGMSKNLARHADKVLQAFDGAFDGDKKVLTVGNPVRQSIANTAPPKRYLNDSDPL 216
Qy 185 RVLVVGGSGOARILNOTMPQVAAKLGDS-VLIWHOSKGSQSQVEQAYAEA---GQPOH 239
Db 217 KVLVGGSLGAKAINEAVVEL-LKLSDKPLTVRHQCGKDNHNTMLVAYSOAQIDTGRHVF 275
Qy 240 KVTFFIDDMAAYAWADVVCRSALTYSIAAAGLPALFPVPOHK-DROQYVNALPLEK 298
Db 276 EVMPFIDDMAQAYSWADVVICRAGALTIVTEIASVGVAAIFVPLPHAVDDHOTANAKSLTD 335
Qy 299 AGAAKITEQPLSDVAVANTLAGWSRETLTMAERARAASIPDATERVANEVSRVARA 356
Db 336 KEAGILLPOHLSGETTLAQILTGLDRQKCLDMAKKAQES-----AKHHVAKTVAEVLVA 389

RESULT 7
US-09-328-352-5036
; Sequence 5036, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5036
; LENGTH: 366

; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5036

Query Match 36.2%; Score 680.5; DB 4; Length 366;
Best Local Similarity 44.3%; Pred. No. 3.2e-62;
Matches 159; Conservative 59; Mismatches 126; Indels 15; Gaps 7;

Qy 7 KRLMVWAGTGGHVPFGLAVAHHLMAQGVWLTADRMADLPVKGIGIEDIRISGL 66
Db 11 KHVMMWAGTGGHVPFGLAVAHHLMAQGVWLTADRMADLPVKGIGIEDIRISGL 70
Qy 67 RGKICALIAPLRIFNARQARIMKAYKPDVVLGMGGYVSGPGGLAAWSLGLIPVVLHE 126
Db 71 RGVVIRKLAAPFKILKATFSAMRYKQLKVDVAGVGGYVAGGGLAARLLGIPVLIHE 130
Qy 127 QNGIAGLTNKLARIATKYMQAEPGAPFNAEV---GNPVRTDVLALPLPQORLAGRE-- 181
Db 131 QNAVAGFTNAQLSRVAKVCEAFENTPASEKVVTTGNPVRRREITDLSPKRYDEREQ 190
Qy 182 -GPVRLVVGSGOARILNOTMPQVAAKLGDSVLIWHOSKGSQSQVEQAYAEAAGQPOH- 239
Db 191 GKPLNLIIVGSLGAKALNERLPALQLEVLNIFHCQCGQVQVEATQALYADA--PAIL 248
Qy 240 --KVTFFIDDMAAYAWADVVCRSALTYSIAAAGLPALFPVPOHK-DROQYVNALPL 296
Db 249 TVQVLPFIEDMAKAYSEADLIICRAGALTIVTEVATAGVAAVFPVLPITAVDDHOTANAKFL 308
Qy 297 EKAGAAKITEQPLSDVAVANTLAGWSRETLTMAERARAASIPDATERVANEVSRV 353
Db 309 ADVGAAKICQOSTWIPE-VLNQLFTLIMNQLLTEMAYKARQHAQPNATGHVVDLIQKM 366

RESULT 8
US-09-984-618-13
; Sequence 13, Application US/08984618
; Patent No. 6251647
; GENERAL INFORMATION:
; APPLICANT: de Lencastre, Herminia
; APPLICANT: Tomasz, Alexander
; TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF
; TITLE OF INVENTION: METHICILLIN RESISTANT BACTERIA AND ANTAGONISTS THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/984,618
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-213 CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1694
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 363 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

Db 9 KILVTGGTGGHIYPALSFVEHVKEAPATBFLYVGTENGLESQIVPKKIPFKTIKIQG 68
 Qy 66 ----LRGKIGIKALIAAPLRFNARQARAIKAYKPDVVLGMGGVSGPGGLAAMSLGIP 121
 Db 69 FKRSLSQPNFKTIYL----FLTSINKAKKIIREFQPDVVGITGGVSGAVVYAAHQLKIP 124
 Qy 122 VYLHEQNGIAGLTNKLARIATKVMQABPGA---PPNAEVV--GNPVRTDVLALPLPQOR 176
 Db 125 TIIHQNSIPGNTNKLARIATKVMQABPGA---PPNAEVV--GNPVRTDVLALPLPQOR 182
 Qy 177 LAGREG--PVR--VLVVGSGQARILNQTMPQVAAKLG--SVIIHQSGKGSQSQSVQOA- 230
 Db 183 ILSEFGLDPAKTVVLFSGSRGALKINQAFQAPLFEEREYQVLYASGERYYQELQESL 242
 Qy 231 -YAEAGQPQHKVTEFIDDMAAYANADVVCGSGALTVSEIAAAGLPAFVFPQOH-KDRQ 288
 Db 243 KSEKKLTNISVQPYIDKMWVNMANTDLMVGRAGATSIAEFTALGLPALIPSPYVINDH 302
 Qy 289 QYWNALPLEKAGAKKIEQPLS-----VDVANTLAGMSRETLTMAERARAASIPDATE 344
 Db 303 QTKNAQSLVKVGAVEMIPDAELTGARLVAIDILL--NNEKQQAATASKGEGIPDASD 360
 Qy 345 RV 346
 Db 361 RL 362

RESULT 11
 US-08-984-618-16
 ; Sequence 16, Application US/08984618
 ; Patent No. 6251647
 ; GENERAL INFORMATION:
 ; APPLICANT: de Lencastre, Herminia
 ; APPLICANT: Tomas, Alexander
 ; TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF
 ; TITLE OF INVENTION: METHICILLIN RESISTANT BACTERIA AND ANTAGONISTS THEREOF
 ; NUMBER OF SEQUENCES: 17
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Klauber & Jackson
 ; STREET: 411 Hackensack Avenue, 4th Floor
 ; CITY: Hackensack
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07601
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/984,618
 ; FILING DATE:
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Jackson Esq., David A.
 ; REGISTRATION NUMBER: 26,742
 ; REFERENCE/DOCKET NUMBER: 600-1-213 CIP
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 201-487-5800
 ; TELEFAX: 201-343-1684
 ; TELEX: 133521
 ; INFORMATION FOR SEQ ID NO: 16:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 363 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHEICAL: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Enterococcus faecalis
 ; US-08-984-618-16

Query Match 21.6%; Score 406.5; DB 3; Length 363;
 Best Local Similarity 30.9%; Pred. No. 9.4e-34;
 Matches 112; Conservative 75; Mismatches 144; Indels 31; Gaps 13;
 Qy 8 RLMVMAGGTGGHVFPGLAVAHLM--AQGWQVRWLTADRMEDLVPKHGIEDFIRISG 65
 Db 2 KILVTGGTGGHIYPALSFVEHVKEAPATBFLYVGTENGLESQIVPKKIPFKTIKIQG 61
 Qy 66 ----LRGKIGIKALIAAPLRFNARQARAIKAYKPDVVLGMGGVSGPGGLAAMSLGIP 121
 Db 62 FKRSLSQPNFKTIYL----FLTSINKAKKIIREFQPDVVGITGGVSGAVVYAAHQLKIP 117
 Qy 122 VYLHEQNGIAGLTNKLARIATKVMQABPGA---PPNAEVV--GNPVRTDVLALPLPQOR 176
 Db 118 TIIHQNSIPGNTNKLARIATKVMQABPGA---PPNAEVV--GNPVRTDVLALPLPQOR 175
 Qy 177 LAGREG--PVR--VLVVGSGQARILNQTMPQVAAKLG--SVIIHQSGKGSQSQSVQOA- 230
 Db 176 ILSEFGLDPAKTVVLFSGSRGALKINQAFQAPLFEEREYQVLYASGERYYQELQESL 235
 Qy 231 -YAEAGQPQHKVTEFIDDMAAYANADVVCGSGALTVSEIAAAGLPAFVFPQOH-KDRQ 288
 Db 236 KSEKKLTNISVQPYIDKMWVNMANTDLMVGRAGATSIAEFTALGLPALIPSPYVINDH 295
 Qy 289 QYWNALPLEKAGAKKIEQPLS-----VDVANTLAGMSRETLTMAERARAASIPDATE 344
 Db 296 QTKNAQSLVKVGAVEMIPDAELTGARLVAIDILL--NNEKQQAATASKGEGIPDASD 353
 Qy 345 RV 346
 Db 354 RL 355
 RESULT 12
 US-09-198-452A-977
 ; Sequence 977, Application US/09198452A
 ; Patent No. 6559294
 ; GENERAL INFORMATION:
 ; APPLICANT: Griffiths, R.
 ; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
 ; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
 ; TITLE OF INVENTION: and treatment of infection
 ; FILE REFERENCE: 9710-003-999
 ; CURRENT APPLICATION NUMBER: US/09/198,452A
 ; CURRENT FILING DATE: 1998-11-24
 ; NUMBER OF SEQ ID NOS: 6849
 ; SEQ ID NO 977
 ; LENGTH: 357
 ; TYPE: PRT
 ; ORGANISM: Chlamydia pneumoniae
 ; US-09-198-452A-977
 Query Match 15.1%; Score 284.5; DB 4; Length 357;
 Best Local Similarity 29.5%; Pred. No. 4.4e-21;
 Matches 104; Conservative 59; Mismatches 142; Indels 47; Gaps 14;
 Qy 7 KRLVMAGGTGGHVFPGLAVAHLM--AQGWQVRWLTADRMEDLVPKHGIEDFIRISGL 66
 Db 6 RKVALVGGSGGHIYPALSFVEHVKEAPATBFLYVGTENGLESQIVPKKIPFKTIKIQG 62
 Qy 67 RG--KGKIKALIAAPLRFNARQARAIKAYKPDVVLGMGGVSGPGGLAAMSLGIPVVL 124
 Db 63 PTVLNPIK-IMSRTLSCSGYLKARKELKIFDDVLIGFSGYHSLPVLVLLAGLSHKIPDLFL 121
 Qy 125 HEQNGIAGLTNKLARIATKVMQABPGA---PPNAEVV--GNPVRTDVLALPLPQOR 174
 Db 122 HEQNLVFGKYNQLFSRYARGIVGNFSPVTKHFRCP-----ABEVFLPKES--FSLGSPM 173
 Qy 175 ORLAGREGPVRVLYVVGSGQARILNQTMPQVAAKLG--SVIIHQSG--KGSQSQSVQOA 230
 Db 174 MKRCTNHTPT-ICVVGSGQAQILNCTVPOALVKLVNKNYPNMYVHHIVGPKSUMKQHV 232
 Qy 231 YAEAGQPQHKVTEFIDDMAAYANADVVCGSGALTVSEIAAAGLPAFVFPQOH-KDRQ 290

Db 233 Y-NRGEVLCVCKPFEQQLLDVLAADLVISRAGATILEILWAKVGGIILIPFGAYGHQE 291
Qy 291 WNA-----LPEKAGAAK-IIIEOPQLSVDA-----VANTLAGMSRE 325
Db 292 VNAKFFVDVLEGGTMIKELTEKLVKVTALDSDHNEKQNSLAAYSQQ 343
RESULT 13
US-08-751-474-2
; Sequence 2, Application US/08751474
; Patent No. 5821335
; GENERAL INFORMATION:
; APPLICANT: Hoskins, JoAnn
; APPLICANT: Skatrud, Paul L.
; TITLE OF INVENTION: Biosynthetic Gene murG From
; TITLE OF INVENTION: Streptococcus Pneumoniae
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/751,474
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-9902
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3334
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 352 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-751-474-2
Query Match 14.5%; Score 273.5; DB 2; Length 352;
Best Local Similarity 27.7%; Pred. No. 5.9e-20;
Matches 93; Conservative 52; Mismatches 150; Indels 41; Gaps 8;
Qy 7 KRLVMVAGGTGGHVPFGLAVAHILMAQGWVRLGTADRMEDLVPKHGIEIDFIRISGL 66
Db 2 KXIVFTGGTGVHTVTLNLLMKPFIEDGNEVHYIGDKRGIEHQEILKSLGLDVPHSA-- 59
Qy 67 RGKGIKALIAAPLRFINAWR-----QARAIMKAYKPDVVLGMGVVSGPGGL 113
Db 60 TGG-----LRRYFSNQNMLDVFKVGMGIVQSLFIMLRPQTLFSKGGFVSPFVI 110
Qy 114 AAWSLGIPVVLHEQNGIAGLTNKLARATKVMQA--EPGAFNAEVVGNPVRDVLALP 171
Db 111 AARVSGVFPVTHESDLSGLANKATYKATKMYSTFEQASSLSKVEHVGATKVSQDNKP 170
Qy 172 LPQ-----QLRAGRGPPVRLVVGSGQARILNQTMPQVAAKLGDSVLIHQSGKGSQCS 226
Db 171 EDELVDIQTENHKLPT-VLFVGGSGARVFNQLVTDHKKELTERYNIINLTGDSLIN- 228
Qy 227 EQVAYEAGQPOHKVTEFIDNMAAYAWADVVCVRSGLTVEISEAAAGLPALFVPFHQKD 286
Db 229 -----ELSQNLFVDYDTLQYPLMLADVVVTGGATIFELLAIKLVHIVPLGREA 292
Qy 287 RQ-QYWNALPLEKAGAAKIIIEOPQLSVDAVANTLA 320

Db 283 SRGQIENNAFYVKGYAEDLQESDULDSLEKLS 318
RESULT 14
US-09-134-001C-3198
; Sequence 3198, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3198
; LENGTH: 366
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3198
Query Match 13.9%; Score 261; DB 4; Length 366;
Best Local Similarity 28.1%; Pred. No. 1.3e-18;
Matches 92; Conservative 55; Mismatches 140; Indels 40; Gaps 11;
Qy 8 RLMVMAGTGGHVPFGLAVAHILMAQGWVRLGTADRMEDLVPKHGIEIDFIRISGLR 67
Db 12 KIATYGGGTGVHVSVNLSTIETKGEAFYIGSKGIEREMIESQLPDIOYVPIS-- 69
Qy 68 GGIKAL-----IAPLRIFNARQARAIMKAYKPDVVLGMGVVSGPGGLAAWSLIGIPW 123
Db 70 GKRRILSFENAKVDFVRLKGLDARKILKKQKEDLFSKGGFVSPVVIARSKIPTI 129
Qy 124 LHEQNGIAGLTNKLARATKVMQAEPGAF-----PNAEVVGNPVRDVLALPLPQQLA 178
Db 130 IHESDLTPGLANKISLKFATKIYTTFTDTLTVLPKDKADFVGATVREDL-----KQG 181
Qy 179 GREGPVR-----VLVVGSGGARILNQTMPQ-VAAKLGDSVLIHQSGKGSQCSV 227
Db 182 NKEGYQLTDFDNKKVLLVVGSGSLGSKLNNIIRQNTALLHDYQII-HLTGK--LV 237
Qy 228 EQVAYEAGQPOHKVTEFIDNMAAYAWADVVCVRSGLTVEISEAAAGLPALFVP--FQH 284
Db 238 DSIINKKGVQF---EFVKDDLTDLAITDTVVSRAAGNAIYEFLSLRIPMLLIPLGLDQ 294
Qy 285 KDRQQYWNALPLEKAGAAKIIIEOPQLS 311
Db 295 SRGQDIDNAKNFESKGYGRHIPEDQLT 321
RESULT 15
US-08-984-618-17
; Sequence 17, Application US/08984618
; Patent No. 6251647
; GENERAL INFORMATION:
; APPLICANT: de Lencastre, Herminia
; APPLICANT: Tomas, Alexander
; TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF
; TITLE OF INVENTION: METHICILLIN RESISTANT BACTERIA AND ANTAGONISTS THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/984,618
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-213 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 525 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Cucumis sativus
US-08-984-618-17

Query Match 7.0%; Score 131; DB 3; Length 525;
Best Local Similarity 20.4%; Pred. No. 7e-05;
Matches 66; Conservative 55; Mismatches 121; Indels 82; Gaps 14;
QY 82 FNAWEQARAIKAYKPDVVLGMGVGSGGLAAMSLGIPVVLHQNQI---AGLTNRWL 138
DB 224 FIAREVAKGLMK-YRPDIISVH-----FLMQHPVPIRLASKGLLNKIV 266
QY 139 -ARIATKVMQAPGAF-----PNAEVG-----NPVTTVLAFLPLP----- 173
DB 267 FTTVVTLSTCHPTWTFHKLIVTRYCYPSTEVAKRATAGLQPSKLKVLGFLVPRSPVKPIR 326
QY 174 ----CORLAGREGVRVLVWGSQCARILNQTMPQVAAKLGDS-----VIIWHQ 218
DB 327 PKIELRKELGMENLPAVLLMGGEGMGPIETAKALSALYDENHGBPIGQVLVICGNH 386
QY 219 SG-KGSQSQVEQAYAEAGQPHKVTETFDIDMAAAYANADVVCVRSALTVEIAAAGLPA 277
DB 387 KKLAGRLRSID-----WKVPVQVKGFTVMBECCMGACDCIITKAGPGGTIAEAMIRGLPI 440
QY 278 L-----FVPFHQKDSQQYNALPLEKAGAAKIIISOPOLSVDVAVANTLAGW---SRETLITWA 331
DB 441 ILNDYIAGQAGNVPY-----VVENGCGKFSKSPK-----EIANIVAKWFGPKADELLIMS 491
QY 332 ERARAASIPDATERVANEVSRVAR 355
DB 492 QNALRLARPDAVFKIVHDLHELK 515

RESULT 16
US-08-984-618-12
Sequence 12, Application US/08984618
Patent No. 6251647
GENERAL INFORMATION:
APPLICANT: de Lencastre, Herminia
APPLICANT: Tomasz, Alexander
TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF
TITLE OF INVENTION: METHICILLIN RESISTANT BACTERIA AND ANTAGONISTS THEREOF
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA

ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/984,618
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-213 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 382 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus subtilis
US-08-984-618-12
Query Match 6.8%; Score 127; DB 3; Length 382;
Best Local Similarity 18.9%; Pred. No. 0.00011;
Matches 59; Conservative 65; Mismatches 118; Indels 70; Gaps 12;
QY 80 RIFNAM-----ROARAIKAYKPDVVLGMGVGSGGLAAMSLGIP-----V 122
DB 83 RKENIYFMGNKGLGELVDEHQDIIINTFPMIWPPEYRRTGRVPTFNVTDFCLHKI 142
QY 123 VLHEQNGIAGLTNWKARIATKVMQAPGAF--NAEVGNPVRTDVLALPLQORLAGRE 181
DB 143 WVHE-----NVDKYVATDYVKEKLEIGHTPSNVAKITGIPR-----PQFESMPV 189
QY 182 GPVR-----VLVWGSQGA-----RLNQTMPQVAAKLGDSVLIHQSGK 221
DB 190 GPIYKYNLSPNKKVLLIMAGAHGVLNKKELCNLVKDDQVQVWVCGKNTAL----- 243
QY 222 GSQSQVEQAYAEAGQPHKVTETFDIDMAAAYANADVVCVRSALTVEIAAAGLPA 279
DB 244 --XESLSALEAENG-KLKVLYGVVERIDELFRITDCMITKPGGIIITETATAIGVPVILYK 300
QY 280 -VPFHQKDSQQYNALPLEKAGAAKIIIEQQLSVDVAVANTLAGMSRETLITMAERARAAS 338
DB 301 FVPGQEK-----NANFFEDRGAAIVVNRHEEILSVTSLLA--DEDTLHRMKNKINDLH 353
QY 339 IPDATERVANEV 350
DB 354 LANSSEVILEDI 365

RESULT 17
US-08-984-618-12
Sequence 26203, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 28203
LENGTH: 572
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26203

Query Match 6.5%; Score 122.5; DB 4; Length 572;
Best Local Similarity 23.8%; Pred. No. 0.00061;
Matches 93; Conservative 47; Mismatches 154; Indels 97; Gaps 17;

QY 18 GHVPPGLAVAHMAQ---GHOVRLGTADMEADLVKXGIBIDFIRIS-GLRGKGIKA 73
DB 157 GYVLADLEQGRFVADLLGGQPVVEGAADWLDAAARLPDPEAALAHVTPSARAPRAE 216
QY 74 LIAAPLRFNWRQARAIMKAYKPDVV-----LGMGYVSGGGL-----AANSLGI 120
DB 217 LLIHPRCVLALEPADAADAARQALVDAGLAAALGGLLADARSADPRLQAAADALGV 276
QY 121 PVVLEHQNGIAGLTNKLARIATKVMQAEFG-----APFNAEVVGNPVRTDVLALPLPQQ 175
DB 277 PLRF-----LPOADPAARLRQALPRAEQVD--RLAVAAAPLAMD 314
QY 176 RLACREGVRVVLVVGSGQARILNQTWPQVAAKLGDSVIIHQSGKGSQSQVEQAYA-EA 234
DB 315 RIGORRG--RLTVIGLGGCAELMVPAPARRALDEADIL-----GYQTYVDMAGFPRA 365
QY 235 GQPKHVTFFDDMAAVAW-----ADVVCVRSGALTVEISIAAGLPAFPVPPQHKDRQ 288
DB 366 GOVLHGSNREELQARHAFELATGRVAVLVVSGDPCVFAMAAVLEAL-----HGAGE 420
QY 289 QYVNALPLE-----KAGAKIIEQP---QLSVDVANTLAGMSRRETLTMAE 332
DB 421 PANDAVELRIFPGVSAALATAARAGA-----PLGHDFCVLSLSDNLKPWE-----VIER 469
QY 333 RARASIPDATERVANEVSR-----VARALE 358
DB 470 RDLAGAADLAWAFNPISRARPWOLGRALE 500

RESULT 18
US-08-311-731A-7
Sequence 7, Application US/08311731A
Patent No. 6583266
GENERAL INFORMATION:
APPLICANT: SMITH, DOUGLAS
APPLICANT: MAO, JEN-I
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 411
CORRESPONDENCE ADDRESS:
ADDRESSER: WOLF, GREENFIELD & SACKS, P.C.
STREET: 600 ATLANTIC AVENUE
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: USA
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311,731A
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: GATES, EDWARD R.
REGISTRATION NUMBER: 31,616
REFERENCE/DOCKET NUMBER: C0044/7125
TELECOMMUNICATION INFORMATION:

TELEPHONE: 617/720-3500
TELEFAX: 617/720-2441
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: MYCOBACTERIUM TUBERCULOSIS
US-08-311-731A-7

Query Match 6.2%; Score 117; DB 4; Length 463;
Best Local Similarity 27.9%; Pred. No. 0.0016;
Matches 46; Conservative 28; Mismatches 51; Indels 40; Gaps 7;

QY 203 POVAAKLGDSVIIHQSGKGSQSQVEQAYAE-----AGQPOHK-----VFTEFD 246
DB 267 PIIVATLGSS-----GKNLLQVVLNALADLPVTVIATAGRNHLKNVPANAFVADYLP 320
QY 247 DMAAAYANADVVCVRSGALTVEISIAAGLPAFPVPPQHKDRQYVWALPLEKAGAKIIE 306
DB 321 GEAAAAARLA-VVLCNGGSGPTTQCALAAGVPVIGLP---SNMDQHLNMEALERAGAGVILLR 376
QY 307 QPQLSVDVANTLAGMSRRETLTMAERARAA-----SIPDA 342
DB 377 TERLNTGVAARV-----KQVLSGAEFRQARRLPKPSDQILPDS 416

RESULT 19
US-09-036-987A-8
Sequence 8, Application US/09036987A
Patent No. 6143526
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H.
APPLICANT: Broughton, Mary C.
APPLICANT: Crawford, Kathryn P.
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Merlo, Donald J.
APPLICANT: Treadway, Patti J.
APPLICANT: Turner, Jan R.
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
TITLE OF INVENTION: Production
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSER: Dow Agrosciences LLC Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/036,987A
FILING DATE: 09-MAR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stuart, Donald R.
REGISTRATION NUMBER: 28,479
REFERENCE/DOCKET NUMBER: 50,608
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317)337-4816
TELEFAX: (317)337-4847
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 390 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-036-987A-8

Query Match
Best Local Similarity 20.6%; Score 115.5; DB 3; Length 390;
Matches 85; Conservative 46; Mismatches 126; Indels 155; Gaps 20;

QY 14 GGTGGHV---FPGLAVAHLLMAQGVQVWLTADRM---EADLVPKHGIEIDFIRISGLR 67
DB 60 GDTGGTTQLRFPNPAFGQRTDTEAGRQL-WEQTASNVAAQSSLDOLPEY-----105
QY 68 GKGKALIAAPLRIFNARQARIMKAYKPDVVLGMGVVSGPGGLAAWSLGIPIVWLH--125
DB 106 -----LRLAEAWRPSVLLV-----DVCALIGRVL-----GGL-----LDLPVVLHRW 142
QY 126 -----EQNGIAGLTNKLWARIATKVMQAEPPGAPFNAEVVGNPVR 164
DB 143 GVDPTAGPFSDBRAHELLDPCVRHGLTGL-----PTPELILDPCP 182
QY 165 TDVLALPLPQORLAGREGPVRVLLVVGGS-----QGARILNQTMP-- 203
DB 183 PSLOASDAQ-----GAPVQVYPYNGSGAFPWGAARTSARRVCICMGRWLNATGPAP 236
QY 204 ---QVAAKL---GDSVIIHQSGKGSQSQVEQAYAEAGOPQHK--VTEFIDDMAAAYA-- 253
DB 237 LIRAVAAATELPGVEAVI-----AVPPEHRLTDLDPNARIAESVP 278
QY 254 -----WADVVCRSALTVEISIAAGLPALFPVPHQKDRQOYWNALPLEKAGAKIIEQ 307
DB 279 LNLFLRTCELVICAGSGTAFTATRLGIPQLVLP-QYFDQFDYARNLAAAGAGICLPDEQ 337
QY 308 POLS-----VDVANTL---AGWSRETLTMAERARAASIPDATERVANEVSRV 353
DB 338 AQSDHEQFTDSIATVLGDTGFA-SAAIKLSDEITAMPHPAALVRLTLENTAAI 388

RESULT 20

US-09-370-700-8
Sequence 8, Application US/09370700
Patent No. 6274350
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H
APPLICANT: Broughton, Mary C
APPLICANT: Crawford, Kathryn P
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Treadway, Patti J
APPLICANT: Turner, Jan E
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
FILE REFERENCE: 50489 DIV1
CURRENT APPLICATION NUMBER: US/09/370,700
CURRENT FILING DATE: 1999-08-09
EARLIER APPLICATION NUMBER: US 09/36987
EARLIER FILING DATE: 1998-03-09
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn ver..2.0
SEQ ID NO 8
LENGTH: 390
TYPE: PRT
ORGANISM: Saccharopolyspora spinosa
US-09-370-700-8

Query Match
Best Local Similarity 20.6%; Score 115.5; DB 3; Length 390;
Matches 85; Conservative 46; Mismatches 126; Indels 155; Gaps 20;

QY 14 GGTGGHV---FPGLAVAHLLMAQGVQVWLTADRM---EADLVPKHGIEIDFIRISGLR 67
DB 60 GDTGGTTQLRFPNPAFGQRTDTEAGRQL-WEQTASNVAAQSSLDOLPEY-----105
QY 68 GKGKALIAAPLRIFNARQARIMKAYKPDVVLGMGVVSGPGGLAAWSLGIPIVWLH--125

Db 106 -----LRLAEAWRPSVLLV-----DVCALIGRVL-----GGL-----LDLPVVLHRW 142
QY 126 -----EQNGIAGLTNKLWARIATKVMQAEPPGAPFNAEVVGNPVR 164
DB 143 GVDPTAGPFSDBRAHELLDPCVRHGLTGL-----PTPELILDPCP 182
QY 165 TDVLALPLPQORLAGREGPVRVLLVVGGS-----QGARILNQTMP-- 203
DB 183 PSLOASDAQ-----GAPVQVYPYNGSGAFPWGAARTSARRVCICMGRWLNATGPAP 236
QY 204 ---QVAAKL---GDSVIIHQSGKGSQSQVEQAYAEAGOPQHK--VTEFIDDMAAAYA-- 253
DB 237 LIRAVAAATELPGVEAVI-----AVPPEHRLTDLDPNARIAESVP 278
QY 254 -----WADVVCRSALTVEISIAAGLPALFPVPHQKDRQOYWNALPLEKAGAKIIEQ 307
DB 279 LNLFLRTCELVICAGSGTAFTATRLGIPQLVLP-QYFDQFDYARNLAAAGAGICLPDEQ 337
QY 308 POLS-----VDVANTL---AGWSRETLTMAERARAASIPDATERVANEVSRV 353
DB 338 AQSDHEQFTDSIATVLGDTGFA-SAAIKLSDEITAMPHPAALVRLTLENTAAI 388

Search completed: June 7, 2004, 07:15:50
Job time : 25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

CM protein - protein search, using sw model

Run on: June 7, 2004, 07:14:21 ; Search time 47 Seconds
(without alignments)

2178.878 Million cell updates/sec

Title: US-09-829-275-1

Perfect score: 1881

Sequence: 1 MYSQGRLLVMAGTGGHV.....RVANEVSRAVALEHHHHH 364

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1155919 seqs, 281338677 residues

Total number of hits satisfying chosen parameters: 1155919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|----------------------|
| 1 | 1881 | 100.0 | 364 | 10 | US-09-829-275-1 |
| 2 | 1803 | 95.9 | 355 | 9 | US-09-741-669-454 |
| 3 | 1803 | 95.9 | 355 | 9 | US-09-815-242-10031 |
| 4 | 1803 | 95.9 | 355 | 12 | US-10-282-122A-56415 |
| 5 | 1687 | 89.7 | 355 | 12 | US-10-282-122A-75189 |
| 6 | 1677 | 89.2 | 355 | 9 | US-09-815-242-14078 |
| 7 | 1677 | 89.2 | 355 | 12 | US-10-282-122A-76021 |
| 8 | 1549.5 | 85.3 | 348 | 12 | US-10-282-122A-59581 |
| 9 | 1549.5 | 82.4 | 348 | 12 | US-10-282-122A-73040 |
| 10 | 1471 | 78.2 | 356 | 12 | US-10-282-122A-77926 |
| 11 | 1345.5 | 71.5 | 360 | 12 | US-10-282-122A-68811 |
| 12 | 1323 | 70.3 | 281 | 12 | US-10-282-122A-55838 |
| 13 | 1154 | 61.4 | 354 | 12 | US-10-282-122A-77456 |
| 14 | 1123 | 59.7 | 354 | 12 | US-10-282-122A-66899 |
| 15 | 1064.5 | 56.6 | 351 | 9 | US-09-815-242-11189 |

| | | | | | | |
|----|--------|------|-----|----|----------------------|--------------------|
| 16 | 1064.5 | 56.6 | 351 | 12 | US-10-282-122A-58422 | Sequence 58422, A |
| 17 | 1061.5 | 56.4 | 351 | 10 | US-09-829-275-2 | Sequence 2, Appli |
| 18 | 871 | 46.3 | 367 | 12 | US-10-282-122A-50225 | Sequence 50225, A |
| 19 | 849 | 45.1 | 372 | 12 | US-10-282-122A-49098 | Sequence 49098, A |
| 20 | 791 | 42.1 | 357 | 12 | US-10-282-122A-51374 | Sequence 51374, A |
| 21 | 772.5 | 41.1 | 355 | 12 | US-10-282-122A-66061 | Sequence 66061, A |
| 22 | 759.5 | 40.4 | 355 | 12 | US-10-282-122A-65125 | Sequence 65125, A |
| 23 | 719.5 | 38.3 | 357 | 9 | US-09-815-242-12009 | Sequence 12009, A |
| 24 | 719.5 | 38.3 | 357 | 12 | US-10-282-122A-66651 | Sequence 66651, A |
| 25 | 717 | 38.1 | 356 | 12 | US-10-282-122A-69709 | Sequence 69709, A |
| 26 | 696.5 | 37.0 | 367 | 12 | US-10-282-122A-63072 | Sequence 63072, A |
| 27 | 684 | 36.4 | 375 | 12 | US-10-282-122A-68322 | Sequence 68322, A |
| 28 | 672.5 | 35.8 | 365 | 12 | US-10-282-122A-44666 | Sequence 44666, A |
| 29 | 535 | 28.4 | 408 | 12 | US-10-282-122A-62025 | Sequence 62025, A |
| 30 | 534 | 28.4 | 250 | 12 | US-10-282-122A-47797 | Sequence 47797, A |
| 31 | 510.5 | 27.1 | 407 | 12 | US-10-282-122A-63788 | Sequence 63788, A |
| 32 | 506 | 26.9 | 410 | 9 | US-09-712-363-223 | Sequence 223, App |
| 33 | 506 | 26.9 | 410 | 12 | US-10-282-122A-62331 | Sequence 62331, A |
| 34 | 506 | 26.9 | 410 | 12 | US-10-282-122A-64551 | Sequence 64551, A |
| 35 | 500 | 26.6 | 410 | 10 | US-09-829-275-8 | Sequence 8, Appli |
| 36 | 488.5 | 26.0 | 359 | 12 | US-10-282-122A-54021 | Sequence 54021, A |
| 37 | 487 | 25.9 | 372 | 9 | US-09-738-626-5863 | Sequence 5863, Ap |
| 38 | 479 | 25.5 | 380 | 12 | US-10-282-122A-48824 | Sequence 48824, A |
| 39 | 470.5 | 25.0 | 409 | 12 | US-10-282-122A-53422 | Sequence 53422, A |
| 40 | 461.5 | 24.5 | 363 | 14 | US-10-156-761-13649 | Sequence 13649, A |
| 41 | 460 | 24.5 | 383 | 12 | US-10-282-122A-45908 | Sequence 45908, A |
| 42 | 423.5 | 22.5 | 366 | 12 | US-10-282-122A-46233 | Sequence 46233, A |
| 43 | 422 | 22.3 | 363 | 12 | US-10-282-122A-60776 | Sequence 60776, A |
| 44 | 420 | 22.3 | 362 | 12 | US-10-282-122A-57874 | Sequence 57874, A |
| 45 | 419 | 22.2 | 360 | 10 | US-09-829-275-4 | Sequence 4, Appli |
| 46 | 418 | 22.2 | 363 | 10 | US-09-829-275-7 | Sequence 7, Appli |
| 47 | 411.5 | 21.9 | 363 | 12 | US-10-282-122A-42564 | Sequence 42564, A |
| 48 | 406.5 | 21.6 | 362 | 10 | US-09-829-275-3 | Sequence 3, Appli |
| 49 | 402.5 | 21.4 | 368 | 9 | US-09-815-242-4866 | Sequence 4866, Ap |
| 50 | 401.5 | 21.3 | 363 | 9 | US-10-282-122A-52804 | Sequence 52804, A |
| 51 | 335.5 | 17.8 | 354 | 12 | US-10-282-122A-51853 | Sequence 51853, A |
| 52 | 328.5 | 17.5 | 359 | 12 | US-10-282-122A-76528 | Sequence 76528, A |
| 53 | 315.5 | 16.8 | 384 | 12 | US-09-829-275-6 | Sequence 6, Appli |
| 54 | 313.5 | 16.7 | 385 | 10 | US-10-282-122A-55084 | Sequence 55084, A |
| 55 | 303 | 16.1 | 357 | 12 | US-10-282-122A-54956 | Sequence 54956, A |
| 56 | 284.5 | 15.1 | 357 | 12 | US-10-282-122A-54956 | Sequence 54956, A |
| 57 | 284.5 | 15.1 | 357 | 15 | US-10-312-273-107 | Sequence 107, App |
| 58 | 284.5 | 15.1 | 357 | 15 | US-10-289-762-977 | Sequence 977, App |
| 59 | 277 | 14.7 | 353 | 9 | US-09-815-242-11564 | Sequence 11564, A |
| 60 | 277 | 14.7 | 353 | 12 | US-10-335-977-7604 | Sequence 7604, Ap |
| 61 | 277 | 14.7 | 363 | 12 | US-10-335-977-7605 | Sequence 7605, Ap |
| 62 | 275.5 | 14.6 | 352 | 9 | US-09-815-242-13634 | Sequence 13634, A |
| 63 | 275 | 14.6 | 353 | 12 | US-10-282-122A-58892 | Sequence 58892, A |
| 64 | 273 | 14.5 | 356 | 9 | US-09-815-242-13614 | Sequence 12614, A |
| 65 | 273 | 14.5 | 356 | 9 | US-09-815-242-12853 | Sequence 12853, A |
| 66 | 273 | 14.5 | 356 | 12 | US-10-282-122A-70334 | Sequence 70334, A |
| 67 | 272.5 | 14.5 | 352 | 10 | US-09-815-242-13399 | Sequence 13399, A |
| 68 | 272.5 | 14.5 | 352 | 10 | US-09-829-275-5 | Sequence 5, Appli |
| 69 | 272.5 | 14.5 | 352 | 12 | US-10-282-122A-73866 | Sequence 73866, A |
| 70 | 263 | 14.0 | 342 | 12 | US-10-282-122A-54490 | Sequence 54490, A |
| 71 | 261 | 13.9 | 357 | 12 | US-10-282-122A-70571 | Sequence 70571, A |
| 72 | 259.5 | 13.8 | 363 | 12 | US-10-282-122A-47259 | Sequence 47259, A |
| 73 | 250 | 13.3 | 360 | 12 | US-10-282-122A-74653 | Sequence 74653, A |
| 74 | 213.5 | 11.4 | 361 | 12 | US-10-282-122A-72041 | Sequence 72041, A |
| 75 | 160.5 | 8.5 | 186 | 12 | US-10-282-122A-71400 | Sequence 71400, A |
| 76 | 149 | 7.9 | 337 | 14 | US-10-156-761-14886 | Sequence 14886, A |
| 77 | 129 | 6.9 | 453 | 12 | US-10-424-599-231702 | Sequence 231702, A |
| 78 | 129 | 6.9 | 464 | 12 | US-10-425-114-55430 | Sequence 55430, A |
| 79 | 122 | 6.5 | 333 | 12 | US-10-425-114-46369 | Sequence 46369, A |
| 80 | 122 | 6.5 | 553 | 12 | US-10-425-114-65133 | Sequence 65133, A |
| 81 | 120 | 6.4 | 411 | 15 | US-10-369-493-562 | Sequence 562, App |
| 82 | 118 | 6.3 | 516 | 12 | US-10-424-599-178191 | Sequence 178191, A |
| 83 | 115.5 | 6.1 | 390 | 16 | US-10-329-148A-8 | Sequence 8, Appli |
| 84 | 113.5 | 6.0 | 437 | 12 | US-10-425-114-39058 | Sequence 39058, A |
| 85 | 113 | 6.0 | 412 | 12 | US-10-425-114-47475 | Sequence 47475, A |
| 86 | 113 | 6.0 | 652 | 14 | US-10-156-761-13141 | Sequence 13141, A |
| 87 | 111 | 5.9 | 389 | 12 | US-10-107-431-223 | Sequence 223, App |
| 88 | 111 | 5.9 | 389 | 12 | US-09-758-759-9 | Sequence 9, Appli |

89 110.5 5.9 727 16 US-10-408-765A-636 Sequence 636, App
90 110.5 5.9 10421 12 US-10-282-122A-61631 Sequence 61631, A
91 109.5 5.8 4471 14 US-10-205-032-10 Appl Sequence 10, Appl
92 108 727 16 US-10-408-765A-1646 Sequence 1646, Ap
93 107.5 5.7 240 12 US-10-424-599-150648 Sequence 150648, A
94 107.5 5.7 412 14 US-10-156-761-8483 Sequence 8483, Ap
95 107 5.7 235 12 US-10-425-114-43991 Sequence 43991, A
96 106.5 5.7 168 14 US-10-360-053-2 Sequence 2, Appl
97 106.5 5.7 168 14 US-10-360-053-2 Sequence 4, Appl
98 106.5 5.7 590 16 US-10-220-480-50 Sequence 50, Appl
99 106.5 5.7 612 16 US-10-220-480-56 Sequence 56, Appl
100 106.5 5.7 1078 14 US-10-156-761-10834 Sequence 10834, A

ALIGNMENTS

RESULT 1
US-09-829-275-1
; Sequence 1, Application US/09829275
; Publication No. US20030077803A1
; GENERAL INFORMATION:
; APPLICANT: WALKER, SUZANNE
; TITLE OF INVENTION: CRYSTALS OF THE ESCHERICHIA COLI MEMBRANE-ASSOCIATED
; TITLE OF INVENTION: GLYCOSYLTRANSFERASE (MURG) PROTEIN, ATOMIC COORDINATES
; TITLE OF INVENTION: AND THREE DIMENSIONAL STRUCTURES THEREOF, ATOMIC
; TITLE OF INVENTION: COORDINATES AND THREE DIMENSIONAL STRUCTURES OF
; TITLE OF INVENTION: BINDING DOMAINS THEREOF, IMAGES THEREOF, AND METHODS
; TITLE OF INVENTION: OF CRYSTALLIZING MURG PROTEINS, MODELS OF UDP-
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES, MURG PROTEINS AND BINDING SITES,
; TITLE OF INVENTION: METHODS OF MAKING MODELS, METHODS OF USING MODELS OF
; TITLE OF INVENTION: MURG, COMPOUNDS THAT BIND, INHIBIT OR STIMULATE MURG
; TITLE OF INVENTION: PROTEINS, AND THERAPEUTIC COMPOSITIONS THEREOF
; FILE REFERENCE: 4555-105
; CURRENT APPLICATION NUMBER: US/09/829,275
; CURRENT FILING DATE: 2001-09-11
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-829-275-1

Query Match 100.0%; Score 1881; DB 10; Length 364;
Best Local Similarity 100.0%; Pred. No. 1.7e-173; Indels 0; Gaps 0;
Matches 364; Conservative 0; Mismatches 0

Qy 1 MMSGQKRLVMVAGGTGGHVFPGLAVAHHLMAQGVQVRLGTADRMADLVPKHGIEIDF 60
Db 1 MMSGQKRLVMVAGGTGGHVFPGLAVAHHLMAQGVQVRLGTADRMADLVPKHGIEIDF 60

Qy 61 IRISGLRGKIGKALIAAPLRIFNARQARIMKAYKPDVVLGNGGYVSGPGGLAANSLGI 120
Db 61 IRISGLRGKIGKALIAAPLRIFNARQARIMKAYKPDVVLGNGGYVSGPGGLAANSLGI 120

Qy 121 PVVLEHQNGIAGLTNKLARIATKVMQAEPCAFNVAEVVGNPVRTDVLALPLPQORLAGR 180
Db 121 PVVLEHQNGIAGLTNKLARIATKVMQAEPCAFNVAEVVGNPVRTDVLALPLPQORLAGR 180

Qy 181 EGPVRVLVVGSGSQGARILNQTMPQVAAKLGDVSVIIHQSGKGSQQSVQVQAYAEAGQPHK 240
Db 181 EGPVRVLVVGSGSQGARILNQTMPQVAAKLGDVSVIIHQSGKGSQQSVQVQAYAEAGQPHK 240

Qy 241 VTEFIDMARAAYADVVVCRSGALTIVSEIAAAGLPALFVFPQHKDQOQYWNALPLEKAG 300
Db 241 VTEFIDMARAAYADVVVCRSGALTIVSEIAAAGLPALFVFPQHKDQOQYWNALPLEKAG 300

Qy 301 AAKTIEQPQLSVDAVANTLAGMSRETLTMAERARAASIPDATERVANEVSRVARALEHH 360
Db 301 AAKTIEQPQLSVDAVANTLAGMSRETLTMAERARAASIPDATERVANEVSRVARALEHH 360

Qy 361 HHHH 364

Db 361 HHHH 364

RESULT 2
US-09-741-669-454
; Sequence 454, Application US/09741669
; Patent No. US20020022718A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; TITLE OF INVENTION: Genes identified as required for
; FILE REFERENCE: ELITRA.009A
; CURRENT APPLICATION NUMBER: US/09/741,669
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 60/173005
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 451
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 454
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-669-454

Query Match 95.9%; Score 1803; DB 9; Length 355;
Best Local Similarity 99.2%; Pred. No. 6.1e-166; Indels 0; Gaps 0;
Matches 352; Conservative 1; Mismatches 2

Qy 2 MMSGQKRLVMVAGGTGGHVFPGLAVAHHLMAQGVQVRLGTADRMADLVPKHGIEIDFI 61
Db 1 MMSGQKRLVMVAGGTGGHVFPGLAVAHHLMAQGVQVRLGTADRMADLVPKHGIEIDFI 60

Qy 62 RISGLRGKIGKALIAAPLRIFNARQARIMKAYKPDVVLGNGGYVSGPGGLAANSLGIP 121
Db 61 RISGLRGKIGKALIAAPLRIFNARQARIMKAYKPDVVLGNGGYVSGPGGLAANSLGIP 120

Qy 122 VVLEHQNGIAGLTNKLARIATKVMQAEPCAFNVAEVVGNPVRTDVLALPLPQORLAGR 181
Db 121 VVLEHQNGIAGLTNKLARIATKVMQAEPCAFNVAEVVGNPVRTDVLALPLPQORLAGR 180

Qy 182 GPVRVLVVGSGSQGARILNQTMPQVAAKLGDVSVIIHQSGKGSQQSVQVQAYAEAGQPHKV 241
Db 181 GPVRVLVVGSGSQGARILNQTMPQVAAKLGDVSVIIHQSGKGSQQSVQVQAYAEAGQPHKV 240

Qy 242 TEFIDMARAAYADVVVCRSGALTIVSEIAAAGLPALFVFPQHKDQOQYWNALPLEKAGA 301
Db 241 TEFIDMARAAYADVVVCRSGALTIVSEIAAAGLPALFVFPQHKDQOQYWNALPLEKAGA 300

Qy 302 AKTIEQPQLSVDAVANTLAGMSRETLTMAERARAASIPDATERVANEVSRVARA 356
Db 301 AKTIEQPQLSVDAVANTLAGMSRETLTMAERARAASIPDATERVANEVSRVARA 355

RESULT 3
US-09-815-242-10031
; Sequence 10031, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242

```
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10031
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10031

Query Match          95.9%; Score 1803; DB 9; Length 355;
Best Local Similarity 99.2%; Pred. No. 6,1e-166;
Matches 352; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 MSGQGKRLVMAGGTGGHVPFGLAVAHLMACQGWVRWLGTDADMEADLVPKHGIEIDFI 61
Db 1 MSGQGKRLVMAGGTGGHVPFGLAVAHLMACQGWVRWLGTDADMEADLVPKHGIEIDFI 60
Qy 62 RISGLRGKIGKALIAAPLRIFNARQARAIKAYKPDVVLGCGVVGSGGLAASLGIP 121
Db 61 RISGLRGKIGKALIAAPLRIFNARQARAIKAYKPDVVLGCGVVGSGGLAASLGIP 120
Qy 122 VVHONGIAGLTNKLARIATKMQAEPGAPPAEAVVGNPVRTDVLALPLPOORLAGRE 181
Db 121 VVHONGIAGLTNKLARIATKMQAEPGAPPAEAVVGNPVRTDVLALPLPOORLAGRE 180
Qy 182 GPRVVLVVGSGQARILNCTMPQVAAKLGDSVTIWHQSGKSGQSQSVQYAEAGQPOHKV 241
Db 181 GPRVVLVVGSGQARILNCTMPQVAAKLGDSVTIWHQSGKSGQSQSVQYAEAGQPOHKV 240
Qy 242 TEFIDMAAAYAWADVVVCRSGALTVSEIAAAGLPALFVPPQHKDRQOQYWNALPLEKAGA 301
Db 241 TEFIDMAAAYAWADVVVCRSGALTVSEIAAAGLPALFVPPQHKDRQOQYWNALPLEKAGA 300
Qy 302 AKIIEQPQLSVDAVANTLAGSRETLLTMAERARAASIPDATERVANEVSRVARA 356
Db 301 AKIIEQPQLSVDAVANTLAGSRETLLTMAERARAASIPDATERVANEVSRVARA 355

RESULT 4
US-10-282-122A-56415
; Sequence 56415, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
```

```
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56415
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-282-122A-56415

Query Match          95.9%; Score 1803; DB 12; Length 355;
Best Local Similarity 99.2%; Pred. No. 5,1e-166;
Matches 352; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 MSGQGKRLVMAGGTGGHVPFGLAVAHLMACQGWVRWLGTDADMEADLVPKHGIEIDFI 61
Db 1 MSGQGKRLVMAGGTGGHVPFGLAVAHLMACQGWVRWLGTDADMEADLVPKHGIEIDFI 60
Qy 62 RISGLRGKIGKALIAAPLRIFNARQARAIKAYKPDVVLGCGVVGSGGLAASLGIP 121
Db 61 RISGLRGKIGKALIAAPLRIFNARQARAIKAYKPDVVLGCGVVGSGGLAASLGIP 120
Qy 122 VVHONGIAGLTNKLARIATKMQAEPGAPPAEAVVGNPVRTDVLALPLPOORLAGRE 181
Db 121 VVHONGIAGLTNKLARIATKMQAEPGAPPAEAVVGNPVRTDVLALPLPOORLAGRE 180
Qy 182 GPRVVLVVGSGQARILNCTMPQVAAKLGDSVTIWHQSGKSGQSQSVQYAEAGQPOHKV 241
Db 181 GPRVVLVVGSGQARILNCTMPQVAAKLGDSVTIWHQSGKSGQSQSVQYAEAGQPOHKV 240
Qy 242 TEFIDMAAAYAWADVVVCRSGALTVSEIAAAGLPALFVPPQHKDRQOQYWNALPLEKAGA 301
Db 241 TEFIDMAAAYAWADVVVCRSGALTVSEIAAAGLPALFVPPQHKDRQOQYWNALPLEKAGA 300
Qy 302 AKIIEQPQLSVDAVANTLAGSRETLLTMAERARAASIPDATERVANEVSRVARA 356
Db 301 AKIIEQPQLSVDAVANTLAGSRETLLTMAERARAASIPDATERVANEVSRVARA 355

RESULT 5
US-10-282-122A-75189
; Sequence 75189, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
```

APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 75189
LENGTH: 355
TYPE: PRT
ORGANISM: Salmonella typhimurium

US-10-282-122A-75189
Query Match 89.7%; Score 1687; DB 12; Length 355;
Best Local Similarity 91.2%; Pred. No. 1.1e-154;
Matches 323; Conservative 18; Mismatches 13; Indels 0; Gaps 0;

QY 2 MSGGKRLMVMAGTGGHVFPGGLAVAHLMQAQGVNWLGTADRMADLVPKHGIDIFI 61
Db 1 MSGGKRLMVMAGTGGHVFPGGLAVAHLMQAQGVNWLGTADRMADLVPKHGIDIFI 60

QY 62 RISGLRGKIGKALLAALPIRIFNARQARAIWKYKPDVVLGMGGYVSGPGGLAANSLGIP 121
Db 61 RISGLRGKIGKALLAALPIRIFNARQARAIWKYKPDVVLGMGGYVSGPGGLAANSLGIP 120

QY 122 VYLHEQNGIAGLTNKLARIATKVMQAEPPGAPFNAEVVGNPVRTDVLALPLPQVRLAGRD 181
Db 121 VYLHEQNGIAGLTNKLAKIATTVMQAFPGAPFNAEVVGNPVRTDVLALPLPQVRLAGRD 180

QY 182 GPRVVLVVGSGGARGILNQTWPQVAALKGDSVITWHQSGKGSQSVQYAEAGQPOHKV 241
Db 181 GPRVVLVVGSGGARGILNQTWPQVAARLGDTVTIWHQSGKGAQLTVEQYAGAGQPOHKV 240

QY 242 TEFIDDMAAAAYAWADVVCRCGALTVEIAAAGLPALFVFPQHKDRQOQYNNALPLEKAGA 301
Db 241 TEFIDDMAAAAYAWADVVCRCGALTVEIAAAGLPALFVFPQHKDRQOQYNNALPLENAGA 300

QY 302 AKIIEQPQLSVDAVANTLAGWSRETLTMAERARAASIPDATERVANEVSRVAR 355
Db 301 AKIIEQPQFTVEAVADTLTLAGWSREALLTMAERARAVSIPDATERVANEVSRVAR 354

RESULT 6
US-09-815-242-14078
Sequence 14078, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel

APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14078
LENGTH: 355
TYPE: PRT
ORGANISM: Salmonella typhi
US-09-815-242-14078

Query Match 89.2%; Score 1677; DB 9; Length 355;
Best Local Similarity 90.7%; Pred. No. 9.9e-154;
Matches 321; Conservative 19; Mismatches 14; Indels 0; Gaps 0;

QY 2 MSGGKRLMVMAGTGGHVFPGGLAVAHLMQAQGVNWLGTADRMADLVPKHGIDIFI 61
Db 1 MSGGKRLMVMAGTGGHVFPGGLAVAHLMQAQGVNWLGTADRMADLVPKHGIDIFI 60

QY 62 RISGLRGKIGKALLAALPIRIFNARQARAIWKYKPDVVLGMGGYVSGPGGLAANSLGIP 121
Db 61 RISGLRGKIGKALLAALPIRIFNARQARAIWKYKPDVVLGMGGYVSGPGGLAANSLGIP 120

QY 122 VYLHEQNGIAGLTNKLARIATKVMQAEPPGAPFNAEVVGNPVRTDVLALPLPQVRLAGRD 181
Db 121 VYLHEQNGIAGLTNKLAKIATTVMQAFPGAPFNAEVVGNPVRTDVLALPLPQVRLAGRD 180

QY 182 GPRVVLVVGSGGARGILNQTWPQVAALKGDSVITWHQSGKGSQSVQYAEAGQPOHKV 241
Db 181 GPRVVLVVGSGGARGILNQTWPQVAARLGDTVTIWHQSGKGAQLTVEQYAGAGQPOHKV 240

QY 242 TEFIDDMAAAAYAWADVVCRCGALTVEIAAAGLPALFVFPQHKDRQOQYNNALPLEKAGA 301
Db 241 TEFIDDMAAAAYAWADVVCRCGALTVEIAAAGLPALFVFPQHKDRQOQYNNALPLENAGA 300

QY 302 AKIIEQPQLSVDAVANTLAGWSRETLTMAERARAASIPDATERVANEVSRVAR 355
Db 301 AKIIEQPQFTVEAVADTLTLAGWSREALLTMAERARAVSIPDATERVANEVSRVAR 354

RESULT 7
US-10-282-122A-76021
Sequence 76021, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Lianguo
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John

APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2001-02-16
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-16
PRIOR APPLICATION NUMBER: 60/269,308
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: Patent in version 3.1
SEQ ID NO 76021
LENGTH: 355
TYPE: PRT
ORGANISM: Salmonella typhi
US-10-282-122A-76021

Query Match 89.2%; Score 1677; DB 12; Length 355;
Best Local Similarity 90.7%; Pred. No. 9.9e-154;
Matches 321; Conservative 19; Mismatches 14; Indels 0; Gaps 0;

Qy 2 MSGQKRLMNVAGTGGHVFGLAVAHLMQAQGVWRLGTADRMADLVPKHGIDIFI 61
Db 1 MSGQKRLMNVAGTGGHVFGLAVAHLMQAQGVWRLGTSDRMADLVPKHGIDIFI 60
Qy 62 RISGLRGKGIKALAAPRIFNARQARA IMKAYKPDVVLGMGGYVSGPGLAANSLGIP 121
Db 61 RISGLRGKGVKALLAAPRIFNARQARA IMKRFKPDVVLGMGGYVSGPGLAANSLGIP 120
Qy 122 VVLHEONGINGLTKWLARIATKMOAEPGAPFNAEVVGNPVRTDVLALPLPOORLAGRE 181
Db 121 VVLHEONGIAGLTQWLAKIATVQAFFGAPFNAEVVGNPVRTDVLALPLQVRLAGRD 180
Qy 182 GPRVVLVVGSGGARILNQTMPQVAALIGDSV IWHQSGKGSQSVQAYAEAGQPQHKV 241
Db 181 GPRVVLVVGSGGARVNLNQTMPQVAALIGDTVTIWHQSGKGAQLTVQAYAGAPQHKV 240
Qy 242 TEFIDDMAAAYADVVYCRSGALTVSEIAAGLPALFVFPQHKDROQYNNALPLEKAGA 301
Db 241 TEFIDDMAAAYADVVYCRSGALTVSEIAAGLPALFVFPQHKDROQYNNALPLENAGA 300
Qy 302 AKIEQPOLSDVAVANTLAGWSRETLTMAERARAASIPDATERVANEVSRRVAR 355
Db 301 AKIEQPQFTVEAVADTLAGWSREALLTMAERARAVSIPDATERVANEVSRRVAR 354

RESULT 8
US-10-282-122A-59581
Sequence 59581, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: Patent in version 3.1
SEQ ID NO 59581
LENGTH: 348
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-10-282-122A-59581

Query Match 85.3%; Score 1604; DB 12; Length 348;
Best Local Similarity 88.8%; Pred. No. 1.1e-146;
Matches 308; Conservative 19; Mismatches 20; Indels 0; Gaps 0;

Qy 10 MYMAGTGGHVPGLAVAHLMQAQGVWRLGTADRMADLVPKHGIDIFIRISGLRGK 69
Db 1 MYMAGTGGHVPGLAVAHLMQAQGVWRLGTADRMADLVPKHGIDIFIRISGLRGK 60
Qy 70 GIKALIAAPRIFNARQARA IMKAYKPDVVLGMGGYVSGPGLAANSLGIPVVLHEONG 129
Db 61 GIKAQALLAPVRIFNARQARA IMKRFKPDVVLGMGGYVSGPGLAANSLGIPVVLHEONG 120
Qy 130 IAGLTNKLARIATKMOAEPGAPFNAEVVGNPVRTDVLALPLPOORLAGREGPVRVLVV 189
Db 121 IAGLTNKLAKIATKVMQAFPGAPFNAEVVGNPVRTDVLALPLQVRLVGRGPIRVLVV 180
Qy 190 GSGGARILNQTMPQVAALIGDSV IWHQSGKGSQSVQAYAEAGQPQHKVTEFIDDMA 249
Db 181 GSGGARVNLNQTMPQVAALIGATVTIWHQSGKGSQQTVCQAAAGAPQHKVTEFIDDMA 240
Qy 250 AAYAWADVVCVRSALTVEIAAGLPALFVFPQHKDROQYNNALPLEKAGAAKILEQPC 309
Db 241 AAYAWADVVCVRSALTVEIAAGLPALFVFPQHKDROQYNNALPLEKAGAAKILEQPC 300
Qy 310 LSVDAVANTLAGWSRETLTMAERARAASIPDATERVANEVSRRVAR 356
Db 301 FTVEAVASTLASWDRETLTMAERARAGASIPDATERVANEVSRAVALA 347

RESULT 9

```
US-10-282-122A-73040
; Sequence 73040, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 73040
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Salmonella paratyphi A
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (259)..(259)
; OTHER INFORMATION: X=any amino acid
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (271)..(271)
; OTHER INFORMATION: X=any amino acid
US-10-282-122A-73040
Query Match 82.4%; Score 1549.5; DB 12; Length 348;
Best Local Similarity 86.5%; Pred. No. 2.2e-141;
Matches 302; Conservative 16; Mismatches 30; Indels 1; Gaps 1;
QY 2 MSGGKELMWAGTGGHVPGLAVAHHLMAQGWVRLGTADRMADLVPRKHGIEIDFI 61
Db 1 MSGGPKLMWAGTGGHVPGLAVAHHLMAQGWVRLGTADRMADLVPRKHGIDIFI 60
QY 62 RISGLRGKIKALIAAPLRFINAWRQARIMKAYKDDVVLGMGGYVSGGGLAANSLGIP 121
Db 61 RISGLRGKIKALIAAPLRFINAWRQARIMKPKFDPVVLGMGGYVSGGGLAANSLGIP 120
QY 122 VVLFHQNGIAGLTNKNLWARTATKMQAEPGAFNAEVVGNFVTDVLAFLPLPQORLAGRE 181
Db 121 VVLFHQNGIAGLTNKNLWARTATVWQAFPGAFNAEVVGNFVTDVLAFLPVAQVLAGRD 180
```

```
US-10-282-122A-77926
; Sequence 77926, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 77926
; LENGTH: 356
; TYPE: PRT
; ORGANISM: Yersinia pestis
US-10-282-122A-77926
Query Match 78.2%; Score 1471; DB 12; Length 356;
Best Local Similarity 78.3%; Pred. No. 9.1e-134;
Matches 278; Conservative 36; Mismatches 41; Indels 0; Gaps 0;
QY 2 MSGGKELMWAGTGGHVPGLAVAHHLMAQGWVRLGTADRMADLVPRKHGIEIDFI 61
Db 1 MSGGPKLMWAGTGGHVPGLAVAHHLMAQGWVRLGTADRMADLVPRKHGIDIFI 60
QY 62 RISGLRGKIKALIAAPLRFINAWRQARIMKAYKDDVVLGMGGYVSGGGLAANSLGIP 121
```

| | | |
|-----|---|-----|
| 61 | KISGLRGKGLMAQLTAPIRIIYAVRQAGKIMRDIQPNVILMGMGYSVGGEGLAANWLUCGP | 120 |
| 122 | VYLHEONGIAGLTNKLWARIATKVMQABPGAFPPNAEVVGNPVRTDLVLPLPQORLAGRE | 181 |
| 121 | VYLHEONGIAGLTNKLWARIATKVMQABPGAFPPNAEVVGNPVRTDLVLPLPQORLAGRE | 180 |
| 182 | GPVRLVWGGSGOARILNQTPWQAAKLGDSVIIWHQSGKSGQSQSVQVQAYABAGQPHKV | 241 |
| 181 | GPVRLVWGGSGOARILNQTPWQAAKLGDSVIIWHQSGKSGQSQSVQVQAYABAGQPHKV | 240 |
| 242 | TEFIDDMAAYAWADVVCRSGLALTYSIIAAGLPAIFVPPQHKDRQOYVNNALPLEKAGA | 301 |
| 241 | TEFIDDMAAYAWADVVCRSGLALTYSIIAAGLPAIFVPPQHKDRQOYVNNALPLEKAGA | 300 |
| 302 | AKTIEOPQLSVDAVANTLAGMSRETLTWAERAAASIPDATERVANEVSRRVA | 356 |
| 301 | AKTIEOPQLSVDAVANTLAGMSRETLTWAERAAASIPDATERVANEVSRRVA | 355 |

RESULT 11
US-10-282-122A-68911
; Sequence 68911, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OR INVENTION: Identification of Essential Genes in Microorganisms

| | | | | |
|-----------------------|--------------|---------------------|----------------|-------------|
| Query Match | 71.5%; | Score 1345.5; | DB 12; | Length 360; |
| Best Local Similarity | 72.1%; | Pred. No. 1.4e-121; | | |
| Matches 259; | Conservative | 40; | Mismatches 55; | Indels 5; |
| | | | | Gaps 1; |

| | | | | | | |
|----|----|--|----------|---------|----------------|-----|
| QY | 61 | MSGQCKLWVWAGTGGGHVPPGLAVAHLLMAQGVQVRLWLTGTA | DMR | EA | DLVPPKHGIEIDFI | 61 |
| DB | 62 | MSRKRRLWVWAGTGGGHVPPGLAVAHLLMAQGVQVRLWLTGTA | DMR | EA | DLVPPKHGIEIEYI | 60 |
| QY | 63 | RISGLRGKGIKALIAAPLRFPNARQAPAIKMYKPDVILGMGGYVSG | PGGLA | AA | NSLGLIP | 121 |
| DB | 64 | RISGLRGKGVKALIAAPIRIKAIKFAQARRIMKYQPDVILGMGGYVSG | PGGVA | AA | NSCGIP | 120 |
| QY | 65 | VYLHEQNGIAGLITNKILARITATKMQAEPGAFNAEVVGNPVRT | DVLA | LP | POORLAGRE | 181 |
| DB | 66 | VYLHEQNGIAGLITNRUSKIAKRVQLQAPGAFANAPVVGNPVRDD | DVLA | LEA | PAERUKGRE | 180 |
| QY | 67 | GPVRLVVGSGQGARILNQTMPQVAALKGDSVIIWHQSGKGSQSV | EQAY | ----- | AEAGQ | 236 |
| DB | 68 | GAVRVLVVGSGQGARILNHTMPVAVGLIGERVTTIWHQAGKGS | SDTKLRYQ | NELS | KNSVK | 240 |
| QY | 69 | POHKUTEIDDMAAYANADVVCESGALTVSETAAAGLPAALFV | PFQKH | QDQRY | YWNALPL | 296 |
| DB | 70 | SEYKYTEFTDDBIAQAYQWADVVCESGALTVSETAAAGLPAI | FVPFQ | KHQDQRY | YWNALPL | 300 |
| QY | 71 | ERAGAAKIIQPOLSDVAVANTLAGWSERTLLTMAERARAAS | TDPAT | ERVANE | SVRVAR | 355 |
| DB | 72 | FWAGRAPIIFONTTPRATATLENDWRDHOIMLMAEKASO | VAITDA | ERVANV | IIIEVAK | 359 |

RESULT 12
US-10-282-122A-55838
; Sequence 55838, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITEA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55838
; LENGTH: 281
; TYPE: PRT
; ORGANISM: Enterobacter cloacae

US-10-282-122A-55838

Query Match 70.3%; Score 1323; DB 12; Length 281;
Best Local Similarity 91.3%; Pred. No. 1.4e-119;
Matches 253; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

QY 5 QKRLWVAGGTGGHVFPGCLAVAHLMAGQWVRWLGTDADMEADLVPKHGIEIDFIRIS 64
DB 3 QPKRLWVAGGTGGHVFPGCLAVAHLMAGQWVRWLGTDADMEADLVPKHGIEIDFIRIS 62
QY 65 GLRGKIGIKALIAAPLRIFNARQARAIMKAYKPDVVLGNGGVSGPGGLAAWSLGI PVVL 124
DB 63 GLRGKGLKAMLLAPVRIFNARQARAIMKAYKPDVVLGNGGVSGPGGLAAWSLGI PVVL 122
QY 125 HQNGIAGLTNKLARIATKVMQAEPCAPNAEVGNVTRDVLALPLPQORLAGREGPV 184
DB 123 HQNGIAGLTNKLARIATKVMQAEPCAPNAEVGNVTRDVLALPLPQORLAGREGPV 182
QY 185 RVLVWGSQGARILNQTMPQVAAKLGDSVIIHQSGKGSQSVQVQAYAGQPQHKVTEF 244
DB 183 RVLVWGSQGARILNQTMPQVAAKLGDSVIIHQSGKGSQSVQVQAYAGQPQHKVTEF 242
QY 245 IDMAAAYAWADVVCRSGLTIVSEIAAAGLPALFVP 281
DB 243 IDMAAAYAWADVVCRSGLTIVSEIAAAGLPALFVP 279

RESULT 13

US-10-282-122A-77456
Sequence 77456, Application US/10282122A
Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 77456

LENGTH: 354

TYPE: PRT
ORGANISM: Vibrio cholerae
US-10-282-122A-77456

Query Match 61.4%; Score 1154; DB 12; Length 354;
Best Local Similarity 64.2%; Pred. No. 4.8e-103;
Matches 224; Conservative 48; Mismatches 77; Indels 0; Gaps 0;

QY 2 MSQSQKRLVMAAGTGGHVFPGCLAVAHLMAGQWVRWLGTDADMEADLVPKHGIEIDFI 61
DB 1 MMNKKLLVMVAGTGGHVFPGCLAVAHLMAGQWVRWLGTDADMEADLVPKHGIEIDFI 60
QY 62 RISGLRGKIGIKALIAAPLRIFNARQARAIMKAYKPDVVLGNGGVSGPGGLAAWSLGI 121
DB 61 QVKGGLRGKIGIKALIAAPLRIFNARQARAIMKAYKPDVVLGNGGVSGPGGLAAWSLGI 120
QY 122 VYLHQNGIAGLTNKLARIATKVMQAEPCAPNAEVGNVTRDVLALPLPQORLAGREG 181
DB 121 VYLHQNGIAGLTNKLARIATKVMQAEPCAPNAEVGNVTRDVLALPLPQORLAGREG 180
QY 182 GPVVRVLVWGSQGARILNQTMPQVAAKLGDSVIIHQSGKGSQSVQVQAYAGQPQHKV 241
DB 181 GAIIRLLVWGSQGARILNQTMPQVAAKLGDSVIIHQSGKGSQSVQVQAYAGQPQHKV 240
QY 242 TEFIDMAAAYAWADVVCRSGLTIVSEIAAAGLPALFVPFQHKDRCQYWNALPLEKAGA 301
DB 241 TEFIDMAAAYAWADVVCRSGLTIVSEIAAAGLPALFVPFQHKDRCQYWNALPLEKAGA 300
QY 302 AKIIEQPOLSVDAVANTLWAGSRETIITLWAEARARAAASIPDATERVANEV 350
DB 301 AKIIEQPOLSVDAVANTLWAGSRETIITLWAEARARAAASIPDATERVANEV 349

RESULT 14

US-10-282-122A-66899
Sequence 66899, Application US/10282122A
Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 77456

LENGTH: 354

PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 58422
LENGTH: 351
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-10-282-122A-50225

Query Match 56.4%; Score 1061.5; DB 10; Length 351;
Best Local Similarity 61.5%; Pred. No. 2.2e-94;
Matches 214; Conservative 44; Mismatches 89; Indels 1; Gaps 1;
QY 5 QGKELVMAGTGGHVFPAIAVAQTLOKQEWDCWLGTKDMEADLPVKIGIDFIRIS 64
DB 2 KKKLLVMAGTGGHVFPAIAVAQTLOKQEWDCWLGTKDMEADLPVKIGIDFIRIS 61
QY 65 GLRGKIGIKALIAAPLRIFNARQARIMKAYKPDVVLGMGGYVSGPGLAAMSLGIPVVL 124
DB 62 GLRGKIGIKALLNAPFAIFRAVLQAKKIIEKPPDAVLMGGYVSGPAGAAKLCGVPIIL 121
QY 125 HEQNGIAGLTINKWLARIATKVMQAEPCGAFNAEVVGNFRTDVLALPQORLAGREGFV 184
DB 122 HEQNAIAGLTINKLGIATCVLQAFPTAFPMAEVVGNFVREDLFEMPNDIRFSDREKL 181
QY 185 RVLVVGSGQARILNQTMPQAAKLGDSVIIHQSGKGSQSQVQAEVQAEAGQPOHKVTEF 244
DB 182 RVLVVGSGQARVNLHTLPKVVAQLADKLEFRHGVGKGAEEVSQLYGE-NLEQVKITEF 240
QY 245 IDDMAAYAVADVVVCRSGALTVSEIAAAGLPALFVFPQHKDQCYWNAFLPXEKAGAAKI 304
DB 241 IDNMAEYAVADVVVCRSGALTVCEIAAVGAAAFVFPQHKDQCYLNKAKYLSVDGAAKI 300
QY 305 IEQPLSDVADVANTLAGWSRETLTMAERARAASIPDATERVANEVSR 352
DB 301 IEQADLTPEILVNVYLNKLTRENLLOMALKAKTMSMFAAQRVAEVIQ 348

RESULT 17

US-09-829-275-2
Sequence 275-2, Application US/09829275
Publication No. US20030077803A1
GENERAL INFORMATION:
APPLICANT: WALKER, SUZANNE
TITLE OF INVENTION: CRYSTALS OF THE ESCHERICHIA COLI MEMBRANE-ASSOCIATED
TITLE OF INVENTION: GLYCOSYLTRANSFERASE (MURG) PROTEIN, ATOMIC COORDINATES
TITLE OF INVENTION: AND THREE DIMENSIONAL STRUCTURES THEREOF, ATOMIC
TITLE OF INVENTION: COORDINATES AND THREE DIMENSIONAL STRUCTURES OF
TITLE OF INVENTION: BINDING DOMAINS THEREOF, IMAGES THEREOF, AND METHODS
TITLE OF INVENTION: OF CRYSTALLIZING MURG PROTEINS, MODELS OF UDP-
TITLE OF INVENTION: GLYCOSYLTRANSFERASES, MURG PROTEINS AND BINDING SITES,
TITLE OF INVENTION: METHODS OF MAKING MODELS, METHODS OF USING MODELS OF
TITLE OF INVENTION: MURG, COMPOUNDS THAT BIND, INHIBIT OR STIMULATE MURG
TITLE OF INVENTION: PROTEINS, AND THERAPEUTIC COMPOSITIONS THEREOF
FILE REFERENCE: 4555-105
CURRENT APPLICATION NUMBER: US/09/829,275
CURRENT FILING DATE: 2001-09-11
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 351
TYPE: PRT
ORGANISM: Haemophilus influenzae
US-09-829-275-2

Query Match 56.4%; Score 1061.5; DB 10; Length 351;
Best Local Similarity 61.5%; Pred. No. 4.3e-94;
Matches 214; Conservative 43; Mismatches 90; Indels 1; Gaps 1;
QY 5 QGKELVMAGTGGHVFPAIAVAQTLOKQEWDCWLGTKDMEADLPVKIGIDFIRIS 64

DB 2 KKKLLVMAGTGGHVFPAIAVAQTLOKQEWDCWLGTKDMEADLPVKIGIDFIRIS 61
QY 65 GLRGKIGIKALIAAPLRIFNARQARIMKAYKPDVVLGMGGYVSGPGLAAMSLGIPVVL 124
DB 62 GLRGKIGIKALLNAPFAIFRAVLQAKKIIEKPPDAVLMGGYVSGPAGAAKLCGVPIIL 121
QY 125 HEQNGIAGLTINKWLARIATKVMQAEPCGAFNAEVVGNFRTDVLALPQORLAGREGFV 184
DB 122 HEQNAIAGLTINKLGIATCVLQAFPTAFPMAEVVGNFVREDLFEMPNDIRFSDREKL 181
QY 185 RVLVVGSGQARILNQTMPQAAKLGDSVIIHQSGKGSQSQVQAEVQAEAGQPOHKVTEF 244
DB 182 RVLVVGSGQARVNLHTLPKVVAQLADKLEFRHGVGKGAEEVSQLYGE-NLEQVKITEF 240
QY 245 IDDMAAYAVADVVVCRSGALTVSEIAAAGLPALFVFPQHKDQCYWNAFLPXEKAGAAKI 304
DB 241 IDNMAEYAVADVVVCRSGALTVCEIAAVGAAAFVFPQHKDQCYLNKAKYLSVDGAAKI 300
QY 305 IEQPLSDVADVANTLAGWSRETLTMAERARAASIPDATERVANEVSR 352
DB 301 IEQADLTPEILVNVYLNKLTRENLLOMALKAKTMSMFAAQRVAEVIQ 348

RESULT 18

US-10-282-122A-50225
Sequence 50225, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Judi
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282.122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 50225
LENGTH: 367
TYPE: PRT
ORGANISM: Burkholderia mallei
US-10-282-122A-50225


```
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51374
; LENGTH: 357
; TYPE: PRT
; ORGANISM: Bordetella pertussis
US-10-282-122A-51374

Query Match          42.1%; Score 791; DB 12; Length 357;
Best Local Similarity 49.3%; Pred. No. 7.3e-68;
Matches 169; Conservative 55; Mismatches 122; Indels 4; Gaps 3;

Qy 9 LMVAGGTGGHVPFGLAVAHHLMAQGVRLGTLADRMEDLVPKHGIIDFIRISGLRG 68
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 6 ILINAGGTGGHINPGLAVAEVLRERGRVLMGNPKMGELVPPRGIELVPLRPQVRG 65
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 69 KGKALIAAPLIRFNARQARINAKYKPDVVLGMGGYVSGPGGLAAWSLGIPTVLHEON 128
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 66 RGAALLKLPFLARACAQAWRLADIRPDVVLGMGGYVAPPGVMAALRRTPLVVHEON 125
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 129 GIAGLTNKLARIATKVMQAEPCGAFNAEVVGNPVRTDVLALPLPOORLAGREGPVRLV 188
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 126 AVAGTANRWLARLARRVLSGFPGLPRGEALGNPVRADLCALPEAERYAGRSGLRLV 185
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 189 VGGSGGARIINQTMPPQVAAKLGDSV--IIWHQSGKSGSQSVQAYAEAGOPQHKVTEPID 246
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 186 VGGSLGAHALNTTVFQALLLPEQARPVVHQAGEQLPALQQAQYAGAG-VQADCRAPID 244
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 247 DMAAAYAWADVVCRSGLTVSEIAAAGLPALFVPFOHK-DROQYWNALPLEKAGAAKII 305
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 245 DMADMAQADLLICRAGANTVSEVAAGCVALLFVFPFPHADDDHQTANAREFLSDAQAAWLQ 304
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 306 EQPQLSVDAVANTLAGWSRETLTMAERARAASIPDATERVANEVSRVAR 355
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 305 PQASLTTPQWLAQWLQRTQBLQAVAGKARTHALPRAAAHIADVCEQAAAR 354
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Search completed: June 7, 2004, 07:20:39
Job time : 49 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 07:01:50 ; Search time 21 Seconds
(without alignments)

1667.320 Million cell updates/sec

Title: US-09-829-275-1

Perfect score: 1881

Sequence: 1 MMSGQKRLVMAGTGGVH.....RVANEVSRLVAREHHHHH 364

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : PIR 78.*

1: Piri.*

2: Piri2.*

3: Piri3.*

4: Piri4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|-----------|--------------------|
| 1 | 1803 | 95.9 | 355 | 1 BVECMG | UDP-N-acetylglucos |
| 2 | 1793 | 95.3 | 355 | 2 F90640 | hypothetical prote |
| 3 | 1793 | 95.3 | 355 | 2 F85491 | hypothetical prote |
| 4 | 1677 | 89.2 | 355 | 2 AE0518 | UDP-N-acetylglucos |
| 5 | 1471 | 78.2 | 356 | 2 AH0068 | UDP-N-acetylglucos |
| 6 | 1154 | 61.4 | 354 | 2 B82091 | UDP-N-acetylglucos |
| 7 | 1054.5 | 56.6 | 351 | 2 D64185 | UDP-N-acetylglucos |
| 8 | 939 | 49.9 | 354 | 2 D84935 | hypothetical prote |
| 9 | 780 | 41.5 | 367 | 2 D82763 | UDP-N-acetylglucos |
| 10 | 779.5 | 41.4 | 355 | 2 A81201 | UDP-N-acetylglucos |
| 11 | 772.5 | 41.1 | 355 | 2 B81777 | UDP-N-acetylglucos |
| 12 | 719.5 | 38.3 | 357 | 2 B83094 | UDP-N-acetylglucos |
| 13 | 510.5 | 27.1 | 407 | 2 D87023 | hypothetical prote |
| 14 | 506 | 26.9 | 410 | 2 E70579 | probable murG prot |
| 15 | 501.5 | 26.7 | 380 | 2 AE3324 | UDP-N-acetylglucos |
| 16 | 484 | 25.7 | 361 | 2 F87565 | hypothetical prote |
| 17 | 470 | 25.0 | 363 | 2 B83970 | UDP-N-acetylglucos |
| 18 | 450.5 | 24.0 | 364 | 2 T34954 | probable UDP-N-ace |
| 19 | 442 | 23.5 | 378 | 2 AG2833 | hypothetical prote |
| 20 | 442 | 23.5 | 378 | 2 C97611 | hypothetical prote |
| 21 | 435 | 23.1 | 357 | 2 B86823 | peptidoglycan synt |
| 22 | 422 | 22.4 | 363 | 2 JCL1275 | phospho-N-acetylm |
| 23 | 422 | 22.4 | 363 | 2 AC1339 | peptidoglycan synt |
| 24 | 419 | 22.3 | 418 | 2 G75496 | UDP-N-acetylglucos |
| 25 | 418 | 22.2 | 363 | 2 AC1700 | peptidoglycan synt |
| 26 | 398.5 | 21.2 | 371 | 2 S76863 | hypothetical prote |
| 27 | 340 | 18.1 | 376 | 2 B97770 | hypothetical prote |
| 28 | 334.5 | 17.8 | 344 | 2 C70401 | phospho-N-acetylm |
| 29 | 332 | 17.7 | 431 | 2 F96784 | hypothetical prote |

| | | | | | | |
|-----|-------|------|------|---|--------|--------------------|
| 30 | 328.5 | 17.5 | 359 | 2 | A97175 | undecaprenyl-PP-Mu |
| 31 | 321.5 | 17.1 | 339 | 2 | B72400 | UDP-N-acetylglucos |
| 32 | 319.5 | 17.0 | 385 | 2 | C71699 | murG protein (murG |
| 33 | 315.5 | 16.8 | 384 | 2 | A71316 | probable UDP-N-glu |
| 34 | 313 | 16.6 | 352 | 2 | H71474 | probable peptidogl |
| 35 | 303 | 16.1 | 353 | 2 | A81737 | UDP-N-acetylglucos |
| 36 | 301.5 | 16.0 | 261 | 2 | A81866 | phospho-N-acetylm |
| 37 | 284.5 | 15.1 | 357 | 2 | F86603 | peptidoglycan tran |
| 38 | 284.5 | 15.1 | 357 | 2 | C72022 | UDP-N-acetylglucos |
| 39 | 277 | 14.7 | 353 | 2 | G71852 | udp-n-acetylglucos |
| 40 | 275 | 14.6 | 353 | 2 | C84664 | transferase, pepti |
| 41 | 273 | 14.5 | 356 | 2 | B89919 | hypothetical prote |
| 42 | 272.5 | 14.5 | 352 | 2 | D79747 | UDP-N-acetylglucos |
| 43 | 269.5 | 14.3 | 352 | 2 | A85080 | hypothetical prote |
| 44 | 263 | 14.0 | 342 | 2 | D81306 | probable UDP-N-ace |
| 45 | 259.5 | 13.8 | 363 | 2 | F70195 | UDP-N-acetylglucos |
| 46 | 162.5 | 8.6 | 388 | 2 | B70878 | probable transfera |
| 47 | 136 | 7.2 | 392 | 2 | S72936 | UDP-glucuronosyltr |
| 48 | 135.5 | 7.2 | 427 | 2 | B95936 | probable glycosylt |
| 49 | 131.5 | 7.0 | 346 | 2 | H95397 | protein imported |
| 50 | 131 | 7.0 | 525 | 2 | T10478 | probable 1,2-diacy |
| 51 | 130.5 | 6.9 | 433 | 2 | S77340 | hypothetical prote |
| 52 | 127 | 6.8 | 382 | 2 | C59935 | cell wall synthesi |
| 53 | 127 | 6.8 | 519 | 2 | H89034 | conserved hypotet |
| 54 | 122.5 | 6.5 | 237 | 2 | A80807 | hypothetical prote |
| 55 | 122.5 | 6.5 | 559 | 2 | F83283 | precorrin-3 methyl |
| 56 | 122 | 6.5 | 383 | 2 | AG2894 | conserved hypotet |
| 57 | 122 | 6.5 | 388 | 2 | B76770 | hypothetical prote |
| 58 | 120 | 6.4 | 411 | 2 | F75439 | probable cell wall |
| 59 | 119.5 | 6.4 | 370 | 2 | B82751 | lipopolysaccharide |
| 60 | 116.5 | 6.2 | 360 | 2 | F95933 | probable glycosylt |
| 61 | 116.5 | 6.2 | 435 | 2 | H86924 | probable glycosyl |
| 62 | 116.5 | 6.2 | 438 | 2 | E86924 | probable glycosyl |
| 63 | 116 | 6.2 | 464 | 2 | C84499 | probable monogalac |
| 64 | 115.5 | 6.1 | 428 | 2 | C70670 | probable glycosyl- |
| 65 | 115.5 | 6.1 | 648 | 1 | P3BPF6 | P3 protein - phage |
| 66 | 115 | 6.1 | 388 | 2 | E72354 | probable hexosyltr |
| 67 | 115 | 6.1 | 428 | 2 | S70670 | 3-deoxy-D-manno-2- |
| 68 | 115 | 6.1 | 449 | 2 | G70670 | hypothetical prote |
| 69 | 115 | 6.1 | 1273 | 2 | E72611 | probable Alf-depen |
| 70 | 114 | 6.1 | 814 | 2 | S65083 | 5-methyltetrahydro |
| 71 | 113 | 6.0 | 533 | 2 | T05092 | probable 1,2-diacy |
| 72 | 110.5 | 5.9 | 727 | 1 | S17854 | NADH2 dehydrogenas |
| 73 | 109.5 | 5.8 | 406 | 2 | A82455 | hypothetical prote |
| 74 | 109 | 5.8 | 392 | 1 | C89851 | macrolide glycosyl |
| 75 | 107.5 | 5.7 | 369 | 2 | H75345 | hypothetical prote |
| 76 | 106.5 | 5.7 | 360 | 2 | E70892 | hypothetical prote |
| 77 | 106.5 | 5.7 | 510 | 2 | G72464 | hypothetical prote |
| 78 | 105.5 | 5.6 | 338 | 2 | C70553 | hypothetical prote |
| 79 | 105.5 | 5.6 | 464 | 2 | F72512 | hypothetical prote |
| 80 | 105.5 | 5.6 | 1111 | 2 | T01239 | hypothetical prote |
| 81 | 105 | 5.6 | 402 | 2 | A87492 | glycosyl transfera |
| 82 | 105 | 5.6 | 458 | 2 | A75386 | conserved hypotet |
| 83 | 105 | 5.6 | 654 | 2 | T34613 | NADH2 dehydrogenas |
| 84 | 104.5 | 5.6 | 537 | 2 | T86883 | hypothetical prote |
| 85 | 104 | 5.5 | 378 | 2 | H70548 | hypothetical prote |
| 86 | 104 | 5.5 | 382 | 2 | T46519 | probable glycosyl |
| 87 | 103.5 | 5.5 | 401 | 2 | T45222 | heat shock protein |
| 88 | 103 | 5.5 | 432 | 2 | T35527 | hypothetical prote |
| 89 | 103 | 5.5 | 1275 | 2 | T85556 | O-antigen biosynth |
| 90 | 102.5 | 5.4 | 395 | 2 | H72654 | hypothetical prote |
| 91 | 102.5 | 5.4 | 395 | 2 | B95938 | hypothetical prote |
| 92 | 102.5 | 5.4 | 404 | 2 | T35421 | probable regulator |
| 93 | 102 | 5.4 | 485 | 2 | T35571 | hypothetical prote |
| 94 | 102 | 5.4 | 510 | 2 | T12970 | hypothetical prote |
| 95 | 102 | 5.4 | 1016 | 2 | T31343 | proline dehydrogen |
| 96 | 101 | 5.4 | 311 | 2 | H87002 | probable dUDP-rham |
| 97 | 101 | 5.4 | 380 | 2 | T35953 | conserved hypotet |
| 98 | 101 | 5.4 | 1616 | 2 | G70668 | polyketide synthas |
| 99 | 100.5 | 5.3 | 370 | 2 | T36352 | hypothetical prote |
| 100 | 100.5 | 5.3 | 534 | 2 | D71344 | hypothetical prote |

ALIGNMENTS

RESULT 1

EVBCMG
UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide)pyrophosphoryl-undecaprenol N-acetyl
C:Species: Escherichia coli
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 01-Mar-2002
C:Accession: JQ0544; JH0093; S40600; B64731
R:Ikeda, M.; Wachi, M.; Jung, H.K.; Ishino, F.; Matsuhashi, M.
Nucleic Acids Res. 18, 4014, 1990
A:Title: Nucleotide sequence involving murG and murC in the mra gene cluster region of E
A:Reference number: JQ0544; MUID:90326550; PMID:2197603
A:Accession: JQ0544
A:Molecule type: DNA
A:Residues: 1-355 <IKS>
A:Cross-references: EMBL:X52644; NID:942053; PIDN:CAA36867.1; PID:942055
R:Mengin-Lecreux, D.; Texier, L.; van Heijenoort, J.
Nucleic Acids Res. 18, 2810, 1990
A:Title: Nucleotide sequence of the cell-envelope murG gene of Escherichia coli.
A:Reference number: JH0093; MUID:90251461; PMID:2187180
A:Accession: JH0093
A:Molecule type: DNA
A:Residues: 1-355 <MEN>
A:Cross-references: EMBL:X52540; NID:942051; PIDN:CAA36776.1; PID:942052
A:Experimental source: strain K12
A:Note: it is uncertain whether Met-1 or Met-9 is the initiator
A:Note: the codon GTA given for residue 274 is inconsistent with the authors' translation
R:Yura, T.; Mori, H.; Nagai, H.; Nagata, T.; Ishihama, A.; Fujita, N.; Isono, K.; Mizobu
submitted to the EMBL Data Library, December 1992
A:Description: Systematic sequencing of the Escherichia coli genome: analysis of the 0-2
A:Reference number: S40531
A:Accession: S40600
A:Molecule type: DNA
A:Residues: 1-355 <YUR>
A:Cross-references: EMBL:D10483; NID:9216434; PIDN:BA01355.1; PID:9216504
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: B64731
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-355 <BLAT>
A:Cross-references: GB:A800118; GB:U00036; NID:91786262; PIDN:AAC73201.1; PID:91786278
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: murG
A:Map position: 2 min
C:Function:
A:Description: involved in murein or cell envelope biosynthesis
A:Pathway: peptidoglycan biosynthesis
C:Superfamily: murG protein
C:Keywords: cell division; cell wall; glycosyltransferase; hexosyltransferase; peptidogl

Query Match 95.9%; Score 1803; DB 1; Length 355;

Best Local Similarity 99.2%; Pred. No. 1.4e-125;

Matches 352; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 MSGQKRLVMVAGTGGHVFPGGLAVAHLMAGQVQVRLGTPADRMEDLVPKHGIEIDFI 61

DB 1 MSGQKRLVMVAGTGGHVFPGGLAVAHLMAGQVQVRLGTPADRMEDLVPKHGIEIDFI 60

QY 62 RISGLRGKIGIKALIAAPLRIFNAMQARAIKAYKPDVVLGMGGYVSGPGLAAMSLGIP 121

DB 61 RISGLRGKIGIKALIAAPLRIFNAMQARAIKAYKPDVVLGMGGYVSGPGLAAMSLGIP 120

QY 122 VYLHEQNGIAGLTNKLAKIATKVMQAFPGAPFNAEVVGNPVRTDVLALPLPQORLAGRE 181

DB 121 VYLHEQNGIAGLTNKLAKIATKVMQAFPGAPFNAEVVGNPVRTDVLALPLPQORLAGRE 180

QY 182 GPRVVLVVGSGGARILNQTWPQVAAKLGDSVTIIHQSGKGSQQSVQAYAEAGQPQHKV 241

DB 181 GPRVVLVVGSGGARILNQTWPQVAAKLGDSVTIIHQSGKGSQQSVQAYAEAGQPQHKV 240

QY 242 TEFIDDMAAAYAWADVVCRSALTVSIIAAGLPALFVFPQHKDROQYWNALPLEKAGA 301

DB 241 TEFIDDMAAAYAWADVVCRSALTVSIIAAGLPALFVFPQHKDROQYWNALPLEKAGA 300

QY 302 AKIIEQPQLSVDVANTLAGRSRETLTMAERARAASIPDATERVANEVSVARA 356

DB 301 AKIIEQPQLSVDVANTLAGRSRETLTMAERARAASIPDATERVANEVSVARA 355

RESULT 2

F90640
hypothetical protein ECs0094 [imported] - Escherichia coli (strain O157:H7, substrain RN
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: F90640
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gaawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: F90640
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-355 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA033517.1; PID:913359550; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECs0094
C:Superfamily: murG protein

Query Match 95.3%; Score 1793; DB 2; Length 355;

Best Local Similarity 98.6%; Pred. No. 7.4e-125;

Matches 350; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 MSGQKRLVMVAGTGGHVFPGGLAVAHLMAGQVQVRLGTPADRMEDLVPKHGIEIDFI 61

DB 1 MSGQKRLVMVAGTGGHVFPGGLAVAHLMAGQVQVRLGTPADRMEDLVPKHGIEIDFI 60

QY 62 RISGLRGKIGIKALIAAPLRIFNAMQARAIKAYKPDVVLGMGGYVSGPGLAAMSLGIP 121

DB 61 RISGLRGKIGIKALIAAPLRIFNAMQARAIKAYKPDVVLGMGGYVSGPGLAAMSLGIP 120

QY 122 VYLHEQNGIAGLTNKLAKIATKVMQAFPGAPFNAEVVGNPVRTDVLALPLPQORLAGRE 181

DB 121 VYLHEQNGIAGLTNKLAKIATKVMQAFPGAPFNAEVVGNPVRTDVLALPLPQORLAGRE 180

QY 182 GPRVVLVVGSGGARILNQTWPQVAAKLGDSVTIIHQSGKGSQQSVQAYAEAGQPQHKV 241

DB 181 GPRVVLVVGSGGARILNQTWPQVAAKLGDSVTIIHQSGKGSQQSVQAYAEAGQPQHKV 240

QY 242 TEFIDDMAAAYAWADVVCRSALTVSIIAAGLPALFVFPQHKDROQYWNALPLEKAGA 301

DB 241 TEFIDDMAAAYAWADVVCRSALTVSIIAAGLPALFVFPQHKDROQYWNALPLEKAGA 300

QY 302 AKIIEQPQLSVDVANTLAGRSRETLTMAERARAASIPDATERVANEVSVARA 356

DB 301 AKIIEQPQLSVDVANTLAGRSRETLTMAERARAASIPDATERVANEVSVARA 355

RESULT 3

F85491
hypothetical protein murG [imported] - Escherichia coli (strain O157:H7, substrain EDL933;
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: F85491
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew,
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: P85491

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-355 <STO>

A:Cross-references: GB:AE005174; NID:g12512793; PIDN:AAG54394.1; GSPDB:GN00145; UWGP:201

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: murG

C:Superfamily: murG protein

Query Match 95.3%; Score:1793; DB 2; Length 355;

Best Local Similarity 98.6%; Pred. No. 7.4e-125;

Matches 350; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 MSGQKRLVMWAGGTGGHVPGLAVAHLMQAQGWQVRLGTADRMEADLVKPKGIEIDFI 61

DB 1 MSAQKRLVMWAGGTGGHVPGLAVAHLMQAQGWQVRLGTADRMEADLVKPKGIEIDFI 60

QY 62 RISGLRGKIKALIAAPLIRIFNAWRQARAIMKAYKPDVVLGMGGYVSGPGGLAAWSLIGIP 121

DB 61 RISGLRGKIKALIAAPLIRIFNAWRQARAIMKAYKPDVVLGMGGYVSGPGGLAAWSLIGIP 120

QY 122 VVLEHQNGIAGLTNKLARIATKMQAEPGAFNAEVVGNPVRTDVLALPLPQORLAGRE 181

DB 121 VVLEHQNGIAGLTNKLAKIATKMQAPPGAFNAEVVGNPVRTDVLALPLPQORLAGRE 180

QY 182 GPRVRLVVGSGQGARILNQTMPQVAAKLGDSVIIHQSGKGSQSQSVQAYAEAGQPQHKV 241

DB 181 GPRVRLVVGSGQGARILNQTMPQVAAKLGDSVIIHQSGKGSQSQSVQAYAEAGQPQHKV 240

QY 242 TEFIDDMAAAAYAWADVVCVRSALTIVSEIAAAGLPALFVFPQHKDRQOQYWNALPLEKAGA 301

DB 241 TEFIDDMAAAAYAWADVVCVRSALTIVSEIAAAGLPALFVFPQHKDRQOQYWNALPLEKAGA 300

QY 302 AKIIEQPQLSDVAVANTLAGRSRETLTMAERARAASIPDATERVANEVSRVARA 356

DB 301 AKIIEQPQLSDVAVANTLAGRSRETLTMAERARAASIPDATERVANEVSRVARA 355

RESULT 4

AB0518 hypotetical protein STY0148 [imported] - Salmonella enterica subsp. enterica serovar Ty

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: AB0518

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;

A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AE0518

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-355 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD01285.1; PID:g16501413; GSPDB:GN00176

C:Genetics:

A:Gene: STY0148

C:Superfamily: murG protein

Query Match 89.2%; Score 1677; DB 2; Length 355;

Best Local Similarity 90.7%; Pred. No. 2.7e-116;

Matches 321; Conservative 19; Mismatches 14; Indels 0; Gaps 0;

QY 2 MSGQKRLVMWAGGTGGHVPGLAVAHLMQAQGWQVRLGTADRMEADLVKPKGIEIDFI 61

DB 1 MSAQKRLVMWAGGTGGHVPGLAVAHLMQAQGWQVRLGTADRMEADLVKPKGIEIDFI 60

QY 62 RISGLRGKIKALIAAPLIRIFNAWRQARAIMKAYKPDVVLGMGGYVSGPGGLAAWSLIGIP 121

DB 61 RISGLRGKIKALIAAPLIRIFNAWRQARAIMKAYKPDVVLGMGGYVSGPGGLAAWSLIGIP 120

RESULT 6

E82081

UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide) pyrophosphoryl-undecaprenol N-ac

C:Species: Vibrio cholerae

QY 122 VVLEHQNGIAGLTNKLARIATKMQAEPGAFNAEVVGNPVRTDVLALPLPQORLAGRE 181

DB 121 VVLEHQNGIAGLTNKLAKIATKMQAPPGAFNAEVVGNPVRTDVLALPLPQORLAGRD 180

QY 182 GPRVRLVVGSGQGARILNQTMPQVAAKLGDSVIIHQSGKGSQSQSVQAYAEAGQPQHKV 241

DB 181 GPRVRLVVGSGQGARILNQTMPQVAAKLGDSVIIHQSGKGAQLTVEQAYAGAGQPQHKV 240

QY 242 TEFIDDMAAAAYAWADVVCVRSALTIVSEIAAAGLPALFVFPQHKDRQOQYWNALPLEKAGA 301

DB 241 TEFIDDMAAAAYAWADVVCVRSALTIVSEIAAAGLPALFVFPQHKDRQOQYWNALPLENAGA 300

QY 302 AKIIEQPQLSDVAVANTLAGRSRETLTMAERARAASIPDATERVANEVSRVAR 355

DB 301 AKIIEQPQLSDVAVANTLAGRSRETLTMAERARAASIPDATERVANEVSRVAR 354

RESULT 5

AH0068

UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide) pyrophosphoryl-undecaprenol N-acet

C:Species: Versinia pestis

C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001

C:Accession: AH0068

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, E

Nature 413, 523-527, 2001

A>Title: Genome sequence of Versinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AH0068

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-356 <KUR>

A:Cross-references: GB:AL590842; PIDN:CAC89411.1; PID:g15978647; GSPDB:GN00175

C:Genetics:

A:Gene: murG

C:Superfamily: murG protein

C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 78.2%; Score 1471; DB 2; Length 356;

Best Local Similarity 78.3%; Pred. No. 4.2e-101;

Matches 278; Conservative 36; Mismatches 41; Indels 0; Gaps 0;

QY 2 MSGQKRLVMWAGGTGGHVPGLAVAHLMQAQGWQVRLGTADRMEADLVKPKGIEIDFI 61

DB 1 MSGTKRLVMWAGGTGGHVPGLAVAHLMQAQGWQVRLGTADRMEADLVKPKGIEIDFI 60

QY 62 RISGLRGKIKALIAAPLIRIFNAWRQARAIMKAYKPDVVLGMGGYVSGPGGLAAWSLIGIP 121

DB 61 KISGLRGKGLMAQLTAPIRIYRAVQAQKIMEDYQPNVVLGMGGYVSGPGGLAAWSLIGIP 120

QY 122 VVLEHQNGIAGLTNKLARIATKMQAEPGAFNAEVVGNPVRTDVLALPLPQORLAGRE 181

DB 121 VVLEHQNGIAGLTNKLARIATKMQAEPGAFNAEVVGNPVRTDVLALPLPQORLAGRE 180

QY 182 GPRVRLVVGSGQGARILNQTMPQVAAKLGDSVIIHQSGKGSQSQSVQAYAEAGQPQHKV 241

DB 181 GPRVRLVVGSGQGARILNQTMPQVAAKLGDSVIIHQSGKGAQLTVEQAYAGAGQPQHKV 240

QY 242 TEFIDDMAAAAYAWADVVCVRSALTIVSEIAAAGLPALFVFPQHKDRQOQYWNALPLEKAGA 301

DB 241 VEFIDDMAAAAYAWADVVCVRSALTIVSEIAAAGLPALFVFPQHKDRQOQYWNALPLEKAGA 300

QY 302 AKIIEQPQLSDVAVANTLAGRSRETLTMAERARAASIPDATERVANEVSRVARA 356

DB 301 AKIIEQPQLSDVAVANTLAGRSRETLTMAERARAASIPDATERVANEVSRVARA 355

C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C/Accession: E82081
R/Heidelberg, J.P.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B.
L. R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 405, 477-483, 2000
A>Title: DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
A/Reference number: A82035; MUID:20406833; PMID:10952301
A/Accession: E82081
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-354 <HEI>
A/Cross-references: GB:AE004310; GB:AE003852; NID:G9656963; PIDN:AAF95544.1; GSPDB:GN001
A/Experimental source: serogroup O1; strain N16961; biotype El Tor
C/Genetics:
A/Map position: 1
C/Superfamily: murG protein

Query Match 61.4%; Score 1154; DB 2; Length 354;
Best Local Similarity 64.2%; Pred. No. 1e-77;
Matches 224; Conservative 48; Mismatches 77; Indels 0; Gaps 0;

QY 2 MSQGGKLMVWAGTGGHVPFGLAVAHLMQAQGVWRWLTADRMADLVPHKGIIDFIRIS 61
DB 1 MNKNKLMVWAGTGGHVPFGLAVAKLQCCQWQIRWLTADRMADLVPHKGIIDFIRIS 60

QY 62 RISGLRKGKIALIAPLRIFNARQBARIMKAYKDDVVLGMGGYVSGPGGLAANSLGIP 121
DB 61 QVXGLRQGLMLKAPFQVNVAILQARRHLITQDPVLMGGYVSGPGGLAANSLGIP 120

QY 122 VILHEQNGIAGLTNKLARIATKVMQAEPCAFNAEVGNVNRDVLALPLPQORLAGREGPV 181
DB 121 VILHEQNAVAGLTNKLAKTAREVQAFGAFADASVGNVNRDVLALPLPQORLAGREGPV 180

QY 182 GPRVVLVWGSQGARILNQTMPQVAAKLGDVSVIWHQSGKSGQSVQEQAYAGQPHKVTEF 241
DB 181 GAIRILVWGSQGARILNQTMPQVAAKLGDVSVIWHQSGKSGQSVQEQAYAGQPHKVTEF 240

QY 242 TERIDDDAAAYADVVVCRSGALTVSEIAAAGLPALFVDFQHKDQVQVWNLPLEKAGAKI 301
DB 241 TERIDDDAAAYADVVVCRSGALTVSEIAAAGLPALFVDFQHKDQVQVWNLPLEKAGAKI 300

QY 302 AKIEQQLSDVADVANTLAGWSRETLTMAERARAASIPDATERVANEV 350
DB 301 AKMIEQQLSDVADVANTLAGWSRETLTMAERARAASIPDATERVANEV 349

RESULT 7
UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide) pyrophosphoryl-undecaprenol N-ace
C/Species: Haemophilus influenzae
C/Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 26-Aug-1999
C/Accession: D64185
R/Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A/Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A./Title: Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.
A/Reference number: A64000; MUID:95350630; PMID:7542800
A/Accession: D64185
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-351 <TIGR>
A/Cross-references: GB:U32793; GB:L42023; NID:G1574683; PIDN:AAC22793.1; PID:G1574693; T
C/Genetics:
A/Genes: murG
C/Function:
A/Description: involved in murin or cell envelope biosynthesis
A/Pathway: peptidoglycan biosynthesis
C/Superfamily: murG protein
C/Keywords: cell division; cell wall; glycosyltransferase; hexosyltransferase; peptidogl

Query Match 56.6%; Score 1064.5; DB 2; Length 351;
Best Local Similarity 61.5%; Pred. No. 4e-71;
Matches 214; Conservative 44; Mismatches 89; Indels 1; Gaps 1;

QY 5 QQKLMVWAGTGGHVPFGLAVAHLMQAQGVWRWLTADRMADLVPHKGIIDFIRIS 64
DB 2 KKKLLVWAGTGGHVPFGLAVAHLMQAQGVWRWLTADRMADLVPHKGIIDFIRIS 61

QY 65 GLRKGKIALIAPLRIFNARQBARIMKAYKDDVVLGMGGYVSGPGGLAANSLGIPVL 124
DB 62 GLRKGKIALIAPLRIFNARQBARIMKAYKDDVVLGMGGYVSGPGGLAANSLGIPVL 121

QY 125 HEQNGIAGLTNKLARIATKVMQAEPCAFNAEVGNVNRDVLALPLPQORLAGREGPV 184
DB 122 HEQNAIAGLTNKLARKATCVLPAPFAFPAFNAEVGNVNRDVLALPLPQORLAGREGPV 181

QY 185 RVLVWGSQGARILNQTMPQVAAKLGDVSVIWHQSGKSGQSVQEQAYAGQPHKVTEF 244
DB 182 RVLVWGSQGARILNQTMPQVAAKLGDVSVIWHQSGKSGQSVQEQAYAGQPHKVTEF 240

QY 245 IDDMAAYADVVVCRSGALTVSEIAAAGLPALFVDFQHKDQVQVWNLPLEKAGAKI 304
DB 241 IDDMAAYADVVVCRSGALTVSEIAAAGLPALFVDFQHKDQVQVWNLPLEKAGAKI 300

QY 305 IEQQLSDVADVANTLAGWSRETLTMAERARAASIPDATERVANEV 352
DB 301 IEQADLTPEILVNVVILKNTRENLQALAKAKTMSMPNAQRAEVIQ 348

RESULT 8
D84955
hypothetical protein murG [imported] - Buchnera sp. (strain APS)
C/Species: Buchnera sp.
C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C/Accession: D84955
R/Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
Nature 407, 81-86, 2000
A/Title: Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. Af
A/Reference number: A84930; MUID:20445173; PMID:10993077
A/Accession: D84955
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-354 <STO>
A/Cross-references: GB:AP000398; GSPDB:GN00144
A/Experimental source: strain APS
C/Genetics:
A/Genes: murG; BU216
C/Superfamily: murG protein

Query Match 49.9%; Score 939; DB 2; Length 354;
Best Local Similarity 48.4%; Pred. No. 7.3e-62;
Matches 169; Conservative 80; Mismatches 100; Indels 0; Gaps 0;

QY 7 KRLVWAGTGGHVPFGLAVAHLMQAQGVWRWLTADRMADLVPHKGIIDFIRISGL 66
DB 4 KKLIIWAGTGGHVPFGLIARVILIEKGLVNWLTGNTKSTIESRIIPKYGKIHYSIKGL 63

QY 67 RKGKIALIAPLRIFNARQBARIMKAYKDDVVLGMGGYVSGPGGLAANSLGIPVLHE 126
DB 64 RNTSLKLLIISPIYLRAYAVAKKIITWSPDVLGMGGYVSGPGGLAANSLGIPVLHE 123

QY 127 ONGIAGLTNKLARIATKVMQAEPCAFNAEVGNVNRDVLALPLPQORLAGREGPV 186
DB 124 QNKTAGITNKLKSLSTKMQAASFGVLRNAEVGNVNRDVLALPLPQORLAGREGPV 183

QY 187 LVWGSQGARILNQTMPQVAAKLGDVSVIWHQSGKSGQSVQEQAYAGQPHKVTEFID 246
DB 184 LVIGSQSSILNRLPEVSPFLKEKIFHWQGTGNYLEKTKKYNKLNQNLITSPK 243

QY 247 DMAAYADVVVCRSGALTVSEIAAAGLPALFVDFQHKDQVQVWNLPLEKAGAKI 306
DB 244 NIAGAYADVVVCRSGALTVSEIAAAGLPALFVDFQHKDQVQVWNLPLEKAGAKI 303

```

307 QPOLSVDAVANTAGWSRETLTVAEERAAASIPDATERVANEVSRRAR 355
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
304 QSNLNTKLIIVNLSDRLDKFLIMAKKAHSLGVRDAIFNFIYVINKISK 352
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
D82763
UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide) pyrophosphoryl-undecaprenol XP07
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: D82763
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: D82763
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-367 <STM>
A:Cross-references: GB:AE003919; GB:AE003949; NID:9105675; PIDN:AAF83607.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Artuda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carrato, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorri, H.; Pacincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; From
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krueger, J.E.; Kuramae, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshakao, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
C:Gene: XF0797
C:Superfamily: murG protein

Query Match 41.5%; Score 780; DB 2; Length 367;
Best Local Similarity 46.9%; Pred. No. 4.1e-50;
Matches 165; Conservative 60; Mismatches 107; Indels 20; Gaps 5;

QY 9 LNMVAGGTGGHVPGLAVAHLMAGQVVRWLGTDADRMEADLVPKHGIEIDFIRISLGR 68
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 15 VMLAGGTGGHVPGLAVAGALRAGVFWMLGATGKMEHLVPHKGIEIQTIAVSGVRG 74

QY 69 KGKALIAAPLRIFNARQARIMKAYKPDVVLGNGGVSGPGLAWSIGIPVVLHQN 128
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 75 HGMALLGTGTVRVLPAIFAAVRVLRRYPRVVRVVFSGPAGGGAARLMLGLPLIVHEQN 134

QY 129 GTAGLTNKLARIATKVMQAEPPNPAEVEVGNVPVRTDVLALPQORLAGREGPVRVLV 188
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 135 RAPGNTNRLARVARVUSGPGSTVAEVEVGNVPVRDIALPAGVRFAGRSFVLLV 194

QY 189 VGGSGCARILNTQMPQAAKLGDSV---SVI IWHQSGKSGSQSQVEQAYAGAQPHKVTFFI 245
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 195 LGSQSGARVMNDALPVLRVLSDADVAVEVRHOCGEALRAETEGAYAGVAA-RISPF 253

QY 246 DMAAAYAWADVVCVRSGALTIVSEITAAAGLPALFVPPQHK-DRQYWNALPLEKAGAKI 304
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 254 SDMAAAYAWADLVCRAGASTIAELCAAGVGSVLPIFFGAVDDHQRRAEYLVVAGALL 313

QY 305 IEQPQLSV-----DAVANTLAGWSRETLTVAEERAAASIPDATERVA 347
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 314 LLQDRAFFVYLVESVLRLLSNFM-----RLAFAEAAARLAKSDVAECTA 359

RESULT 10
A81201
UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide) pyrophosphoryl-undecaprenol N-ac

C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: A81201
R:Petelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ver
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: A81201
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-355 <TE>
A:Cross-references: GB:AE002398; GB:AE002098; NID:G7225640; PIDN:AAF40860.1; PID:G722564;
A:Experimental source: serogroup B, strain MC58
C:Genetics:
C:Gene: NMB0422
C:Superfamily: murG protein

Query Match 41.4%; Score 779.5; DB 2; Length 355;
Best Local Similarity 46.2%; Pred. No. 4.3e-50;
Matches 160; Conservative 73; Mismatches 108; Indels 5; Gaps 4;

QY 6 GKRLVMAGGTGGHVPGLAVAHLMAGQVVRWLGTDADRMEADLVPKHGIEIDFIRISG 65
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 GKTFLMAGGTGGHVPGLAVADSLRARGHVRWLGSKDSMEERIVPQYGRLETLAIG 62

QY 66 LRKGTGKALIAAPLRIFNARQARIMKAYKPDVVLGNGGVSGPGLAAWSLGPVVLH 125
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 VRNGIKRKLMPVLVITQVTRAQRIIRKRVCEVIGGFTVFGGLAAKLGVPVIVH 122

QY 126 EQNGIAGLTNKLARIATKVMQAEPPN-ABVVGNNPVRTDVLALPQORLAGREGPV 184
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 123 EQNAVAGLSNRHLSRWAKVLYAFPKATFSHEGGLVGNPVRADISNLPVPAERFQGREGL 182

QY 185 RVLVVGSGSOGARILNTQMPQAAKLGDSV---IWHQSGKSGSQSQVEQAYAGAQPHKVT 242
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 183 KILWGGSLGADVLANKTVPQALLPDNPOMYQSGRGKLSGLQADYDALG-VKAEVC 241

QY 243 EFIDDMAAAYAWADVVCVRSGALTIVSEITAAAGLPALFVPPQHK-DRQYWNALPLEKAGA 301
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 242 EFTDMVSAVDADLVICRAGALTIAELTAAGLGALLVFPYHVAVDHQTANARFMVQAEA 301

QY 302 AKIEQPQLSVDAVANTLAGWSRETLTVAEERAAASIPDATERVA 347
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 302 GLLLPQTLTAELKLAELGLNREKLRKWAENARTLALPHSADDVA 347

RESULT 11
B81777
UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide) pyrophosphoryl-undecaprenol N-ac

C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: B81777
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel]
; Holroyd, S.; Jørgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: B81777
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-355 <FAR>
A:Cross-references: GB:AL157959; GB:AL162758; NID:G7380672; PIDN:CAB85280.1; PID:G738069(
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
C:Gene: murG; NMA2062
C:Superfamily: murG protein
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 41.1%; Score 772.5; DB 2; Length 355;

```

Best Local Similarity 46.2%; Pred. No. 1.4e-49;
Matches 160; Conservative 71; Mismatches 110; Indels 5; Gaps 4;
QY 6 GKRLMWAGTGGHVFGLAVAHLMQAGQVWLTGADMEADLVPKHGIEIDFIRISG 65
Db 3 GKTFLMAGTGGHVFGLAVADSLRARGHHVWLGSKOSMEIRIVPQDILLETIAIG 62
QY 66 LRKGKIKALIAAPLIRFNARQARIMKAYKPDVVLGMGGYVSGPGGLAAWSLIGIPVLH 125
Db 63 VRNGIKRKLMLPTLTQTVREAQIIRKHREVCVIGFGVTFPGGLAAKLGVPIVH 122
QY 126 BONGIAGLTKWLARIATKVMQAPGAPFN-AEVGNPVRTDVLALPLPOORLAGREGPV 184
Db 123 EQNAVAGLSNRHLRWAKRYLYAPPKAFSEGGVLGNVEVRAAD:SNLFPVPAERFQSGEGL 182
QY 185 RVLVVGSGQARILNQTPQVAAKLGDSV--IIWHQSGKSGQSQVQAYAAQPOQHKVT 242
Db 183 KILVVGSLGADVLTNKVTPQALALLPNAAPQVMYHQSGRKLQSLQADYALG-VQAEV 241
QY 243 EFIDMAAAYAVADVVCRSGLATVSEIAAGLPALFPVFOHK-DROQYVNALPLEKAGA 301
Db 242 EFIDMVSAVYADLVICRAGLTIAELTAAGLGLALLVPYPAHVDDHOTANARFVQAEA 301
QY 302 AKIIEQQLSVDVANTLAGSRETLTMAERARASIPDATERVA 347
Db 302 GLLPQIQTAEKAEILGGLNREKCKWAENARTLALPHSADVA 347
RESULT 12
UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide) pyrophosphoryl-undecaprenol N-ac
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: E83094
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bha
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
: Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: E83094
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-357 <STO>
A:Cross-references: GB:AE004856; GB:AE004091, NID:g9950633; PIDN:AAG07800.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: murG; PA4412
C:Superfamily: murG protein
Query Match 38.3%; Score 719.5; DB 2; Length 357;
Best Local Similarity 44.5%; Pred. No. 1.1e-45;
Matches 158; Conservative 64; Mismatches 122; Indels 11; Gaps 6;
QY 9 LMVWAGTGGHVFGLAVAHLMQAGQVWLTGADMEADLVPKHGIEIDFIRISGLRG 68
Db 5 VLMWAGTGGHVFGLAVAHLMQAGQVWLTGADMEADLVPKHGIEIDFIRISGLRG 64
QY 69 KGKIKALIAAPLIRFNARQARIMKAYKPDVVLGMGGYVSGPGGLAAWSLIGIPVLH 128
Db 65 KGLKSLVKAPELKLKSLFQALVRVIRQRCVCLGGYVTPGGLAARLNGVFLVH 124
QY 129 GIAGLTKWLARIATKVMQAPGAPFN--VVGNPVRTDVLALPLPOORLAGREGPV 185
Db 125 AVAGTANRSLAPIARRVCEAFDFFPASDKLTTGNVPGELFLDAHARAPLTGRR--VN 182
QY 186 VLVVVGSGQARILNQTPQVAAKLGDSV--IIWHQSGKSGQSQVQAYAAQPOQHKVT 243
Db 183 LVLVVGSLGAPFNKLLPEALQAVPLRPIRAIRHQARQHAETIAERTVTA-VEADVAP 241
QY 244 FIDMAAAYAVADVVCRSGLATVSEIAAGLPALFPVFOHK-DROQYVNALPLEKAGA 302
Db 242 FIDMAAAYAVADLVICRAGLTIVSELTAAGLPALFPVHIDHQTNAEFVRSAG 301

QY 303 KIIIEQQLSVDVANTLAG--WSRETLTMAERARASIPDATERVANEVSRVAR 355
Db 302 RLLPKSGTGAELAAQLSEVLHMHPTLRSMAQDQARSLAKPEATRTVVVDACLEVAR 356
RESULT 13
Db7023
hypothetical protein murG [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: D87023
R:Colb, S.T.; Biglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hor
; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21125732; PMID:11234002
A:Accession: D87023
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-407 <STO>
A:Cross-references: GB:AL450380; NID:g13092984; PIDN:CAC31295.1; GSPDB:GN00147
C:Genetics:
A:Gene: murG
C:Superfamily: murG protein
Query Match 27.1%; Score 510.5; DB 2; Length 407;
Best Local Similarity 37.4%; Pred. No. 3.6e-30;
Matches 137; Conservative 62; Mismatches 138; Indels 29; Gaps 9;
QY 9 LMVWAGTGGHVFGLAVAHLMQAGQV--WLTGADMEADLVPKHGIEIDFIRISGL 66
Db 31 VVLWAGTGGHVFGLAVAHLMQAGQV--WLTGADMEADLVPKHGIEIDFIRISGL 90
QY 67 RKGIKALIAAPLIRFNARQARIMKAYKPDVVLGMGGYVSGPGGLAAWSL----GIP 121
Db 91 PKLTGLDLARLPLRVRAVRETRAVFEVAAHVVGFGYVALPAYLAARGIPVRRIP 150
QY 122 VVLHONGIAGLTKWLARIATKVMQAPGGA-FPNAEVGNPVRTDVLALPLPOORLAGR 180
Db 151 VVHHEANARACIANRVGRTAERVLSAVPGSLRGAEVVGVPFIHATITTLNRPALRADAR 210
QY 181 E-----GPVRLVW--GSGQARILNQTPQVAAKLGDS--VIWHQSGKSGQSQVQAYAE 233
Db 211 KHFGTDDARVLLVFGSGQAVSLNRVAGAEADLAASGVAVLH--AYGLKNTLELRTP 268
QY 234 AGQPOHKVTEFIDMAAAYAVADVVCRSGLATVSEIAAGLPALFPVFOHKDQOYVNA 293
Db 269 YGEPYVAVPYLDRLMDLAYAAADLVICSGAMTVAEVSAGVLPAYVFPFPIGNQGLRNA 328
QY 294 LPLEKAGAAKTIEQQLSVDVANTLAGSRETLTMAERARASIPDATERV-----AN 348
Db 329 LPVWAGGLVVDADLTPLGLV-----RQVVLPSDPAQLAATAAARVGHDAH 381
QY 349 EVSRVA 354
Db 382 HVAKVA 387

RESULT 14
E70579
probable murG protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: E70579
R:Colb, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: E70579
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-410 <COL>
A:Cross-references: GB:Z95388; GB:AL123456; NID:93261759; PIDN:CAB08640.1; PID:G2104325
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: murG
C:Superfamily: murG protein

Query Match 26.9%; Score 506; DB 2; Length 410;
Best Local Similarity 37.8%; Pred. No. 7.7e-30;
Matches 143; Conservative 55; Mismatches 148; Indels 32; Gaps 10;

QY 3 SGQKELMVMAGGTGGHVPFGLAVAHLMAGQWQVR--WLGTFADMEADLVPKHGIEIDF 60
DB 32 SADSUSVVLGGGTAGHVEPAMADALVALDPRVITATGLTGLGELVLPQSGYHLEL 91
QY 61 IRISGLRGKIGKIALTAAPLIRIFNAWRQARIMKAYKPDVVLGMGVGSGPGLAANSL-- 118
DB 92 ITAVPMRPFPGDGLARLPSPVRVRAVEARDVLDVADVVVGFYGYVALPAYLAARGLPL 151
QY 119 -----GIPVVLHEQNGIAGLTKWLA---RIATKVMQAEFGAFNAEVV--GNPVRTDVLALPLP 173
DB 152 PPRRRRIRPIVHEANARAGLANRGAHTADVLGSAVDSGLRRRAEVGVGVPVRSIAALD 211
QY 172 LPQORLAGR-----EGPVRVLV--GSGQGARILNQTMPQVAAKGLD--SVIIWHQSGKGSQ 224
DB 212 RAVLRAEAAHFGFPDDARVLLVFGSGQGAVALNRAVSGAALAAAGCVLH--AHGFPQ 269
QY 225 QVQAYAEAGQPOHKVTEFIDDMAAAYAWADVVCRSALTVSSEIAAGLPALFVPPQHK--DRQ 288
DB 270 NVLELRRAAGGPPYVAVPYLDRMELAYAAADLVCRAGAMTVAEVSAGVPAIYVPLPI 329
QY 285 KDRQQVYNALPLEKAGAKIIIEQPOLSDVAVANTL--AGHSRETLTMAEARAASIPDATERV 346
DB 330 GNGEQELNALPVNAGGGMVADAAALTPELVARQVAG-----LLT--DPAALAATAAA 358
QY 345 RV-----ANEVSRVA 354
DB 383 RVGHRDAAGQVAAALV 366

RESULT 15
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: E70579
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-410 <COL>
A:Cross-references: GB:Z95388; GB:AL123456; NID:93261759; PIDN:CAB08640.1; PID:G2104325
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: murG
C:Superfamily: murG protein

Query Match 26.7%; Score 501.5; DB 2; Length 380;
Best Local Similarity 36.7%; Pred. No. 1.5e-29;
Matches 135; Conservative 55; Mismatches 157; Indels 21; Gaps 10;

QY 2 MSQGLVMVAGGTGGHVPFGLAVAHLMAGQWQVRMLGT---ADRMEADLVPKHGIEI 58
DB 32 SADSUSVVLGGGTAGHVEPAMADALVALDPRVITATGLTGLGELVLPQSGYHLEL 91

DB 5 LANQGV--IVLAGGTGCHLFPFAELAHFELRARGWDVH--LATDARAQRFFVGAFADQH---V 59
QY 59 DFRISGLRGKIGKIALTAAPLIRIFNAWRQARIMKAYKPDVVLGMGVGSGPGLAANSL 118
DB 60 HVIRSATIAGNRVVALTKTWSLQGNLDSRKLPRLKPKLVVGFYGYPTLPPLPYAASNM 119
QY 119 GIPVVLHEQNGIAGLTKWLA---RIATKVMQAEFGAFNAEVV--GNPVRTDVLALPLP 173
DB 120 GIPTLIHEQNAVGRANKGLAGRVKAIAGGFLPENSQAATAKTITGNPVRSPVLAAT 179
QY 174 QORLAGREGPVRVLVVGSGQGARILNQTMPQVAAKGLD--SVIIWHQSGKGSQSQVEQ 229
DB 180 PYPAGKDDFRLLVFGSGQGAOFFSOALPAVALLPEHERARLLITQARKDEASARQ 239
QY 230 AYAEAGQPOHKVTEFIDDMAAAYAWADVVCRSALTVSSEIAAGLPALFVPPQHK--DRQ 288
DB 240 AYEKLGVPA--DVAPFFNDMEARMADAHFVIARSASTVSEITVIGRPAMLVPPFHALDHD 298
QY 289 QYVNALEPLEKAGAKIIIEQPOLSDVAVANTL--AGHSRETLTMAEARAASIPDATERV 346
DB 299 QAANAALAAAGAEVVRQADLSQRLAEMLOSAMNELERLEQQAKAASVGVKPDARLL 358
QY 347 ANEVSRVA 354
DB 359 ADLAEATA 366

RESULT 16
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: E70579
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-361 <STO>
A:Cross-references: GB:AE005673; NID:G13424116; PIDN:AAK24522.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC2551
C:Superfamily: murG protein

Query Match 25.7%; Score 484; DB 2; Length 361;
Best Local Similarity 37.2%; Pred. No. 2.7e-28;
Matches 139; Conservative 55; Mismatches 140; Indels 40; Gaps 11;

QY 7 KRLMVAGGTGGHVPFGLAVAHLMAGQWQVRMLGTADR--MEADLVPKHGIEIDFIRIS 64
DB 3 KLAVVAGGTGGHVPFGLAVAHLMAGQWQVRMLGTADR--MEADLVPKHGIEIDFIRIS 64
QY 65 GLRGKIGKAL--IAAPLIRIFNAWRQARIMKAYKPDVVLGMGVGSGPGLAANSLGIPV 122
DB 58 AATAKSNDFLGMKAGFVVLQGVWEARAFKRLDPAVVVGFYGYPALPALLGALSQGRPT 117
QY 123 VLHEQNGIAGLTKWLAIRATKVM-----QAEPCAFNAEVVGNPVRTDVLAL--PLPQ 174
DB 118 VIHEQNAVLRVNRFLAPRVNEVACAPFLEKATPAVKACAHVYGNPVRFPFVRALEFDP- 176
QY 175 QRLAGREGPVRVLVVGSGQGARILNQTMPQVAAKGLDVSIIWHQSGKGSQSQVEQYAE 234
DB 177 --YLAPEVQLRVLTGSGQGARLLSEIPEAVAKLPEM-----RGLKVPQQAAR 227
QY 235 GQPOHK-----VTEFIDDMAAAYAWADVVCRSALTVSSEIAAGLPALFVPPQ- 283
DB 228 MEQARKVYRNAMVECEVAPFPRDMAGYLRQSHLIVGRSGASTCTELAVAGRPISILPLKI 287
QY 284 HKDRQQVYNALPLEKAGAKIIIEQPOLSDVAVANTL--AGHSRETLTMAEARAASIPD 341

Db 288 AADHQRNFARLLEBAGAAVCLELTDVYMAAALKALLSKPERLERKMAAGARSAAKN 347
Qy 342 ATERVANEVSERVAR 335
Db 348 AABELADLVKTKAR 361

RESULT 17
E83970
UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide)pyrophos murG [imported] - Bacill
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: E83970
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: E83970
A:Status: preliminary
A:Molecule type: DNA
A:Cross-references: ST0
A:Experimental source: strain C-125
C:Genetics:
A:Gene: murG
C:Superfamily: murG protein

Query Match 25.0%; Score 470; DB 2; Length 363;
Best Local Similarity 32.5%; Pred. No. 3e-27;
Matches 123; Conservative 78; Mismatches 131; Indels 46; Gaps 12;

Qy 8 RLWVAGTGGHVPGLAVAHHLMA--QGVQVRLWGLTADRMEDLVPKHGIEIDFIRISG 65
Db 2 KIVVSGGTGTHVPALAF.NEMKKRERLDVLYIGTERGLESEIVPREGIPPTIHTG 61

Qy 66 LRGGIKALIAAPLIRIFNARQARIMKAYKPDVVLGGVYSGPGGLAAMSIGIPVHLH 125
Db 62 FQRLSMENKTVVRFLRGTKRAKALLNEFKPDVIGTGGVGGVGPVYAAAKIKITVTH 121

Qy 126 EONGIAGTNNKLA---RIATKMQAEPGAFNAEVW--GNPVRTDVLALPLPQORLAG 179
Db 122 EQNSVPLTNNKFLSYVDRIAICPKAE-AFFPKNVVFTGNPRAEVMG-----GN 172

Qy 180 REGPVR-----VLVVGSGSGARILNQT-----MPQVAALGDSVIRIWHQSGKSGOQ 225
Db 173 REEGLRSLGIKPKKTKVLIVGSGRGARINDAFWSILSDVKAKPYQFVYV---TGTVHYE 229

Qy 226 SVEQAYAEAGOPQHKVTE-FIDDMAAYANADVVCESGALTYSETAAGLPAFLVFPQH 284
Db 230 RVQEQMSIGQFENVIVQPFTHNPVDVLSAVDLIVARAGATTAEITAGLPLSPY 289

Qy 285 -KDRQYVWNLPLEKAGAATIEQPLSDVAVANTLAGWSRETLT-----MAERARAA 337
Db 290 VTNNHQEKNAALSKDAAILKESLITGDRLEDI----DDIWTFTGRDLAMKQAKAL 345

Qy 338 SIPDATERVANEVSERVAR 355
Db 346 GVPTAAEKHLMLREVAK 363

RESULT 18
T34954
probable UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide) pyrophosphoryl-undecapre
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 21-Jan-2000
C:Accession: T34954
R:Saunders, D.C.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1999
A:Reference number: Z21563
A:Accession: T34954
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-364 <SAU>
A:Cross-references: EMBL:AL109663; PIDN:CABS1993.1; GSPDB:GN00070; SCOEDB:SC4A10.17C
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: murG; SCOEDB:SC4A10.17C
C:Superfamily: murG protein

Query Match 24.0%; Score 450.5; DB 2; Length 364;
Best Local Similarity 34.6%; Pred. No. 8.2e-26;
Matches 120; Conservative 58; Mismatches 142; Indels 27; Gaps 7;

Qy 9 LMWVAGTGGHVPGLAVAHHLMAQGHQV--RWLGTADRMEDLVPKHGIEIDFIRISGL 66
Db 3 VVLAGGTAGHIEPALADALREQDPTVGTALGTERTGLTRLVPRKGYELALIPVPL 62

Qy 67 RKGKIKALIAAPLIRIFNARQARIMKAYKPDVVLGGVYSGPGGLAAMSIGIPVHLH 126
Db 63 PRKTPPELITVPGRLRGITKATEQILERTKADAVAGGVYVALPAYLAARLGVPIVH 122

Qy 127 QNGIAGTNNKMLARIATKMQAEPGA-FPNAEVVGNVVRTDVLALPLPQORLAGR----- 180
Db 123 ANARPLANKIGSRVAAQVAVSTPDSKLSRYSYIGIPLRISIATLDRAAARPERAMFGL 182

Qy 181 -EGPVRLVVGSGSGARILNQTMPVAAKLGDSVI-IWHQSGKSGSQSVQAYAEAGOPQ 238
Db 183 DPNLPTLLVTCGSGQARRLENEVIQQVAPWLQQAQIGIQLH--AVGPKNELPQVHQMGPMP 240

Qy 239 HKVTEFIDDMAAAYAWADVVCESGALTYSETAAGLPAFLVFPQHKDROQYVWNLPLEX 298
Db 241 YIPVSYLDRMDLAVAADMLCRAGMTVAELSAVGIPAAVYVPLPIGNGBQELNAQPVVK 300

Qy 299 AGAKIKIQQPOLSDVAVANTLAGWSRETLT-----TMAERARAAAS 338
Db 301 AGGGLLVDDAELTPE-----WLQNVLPVLADPHRLVEMSRAAA 339

RESULT 19
AG2833
hypothetical protein murG [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AG2833
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AG2833
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-378 <KUR>
A:Cross-references: GB:AB008688; PIDN:AAL43085.1; PID:g17740555; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: murG
A:Map position: circular chromosome
C:Superfamily: murG protein

Query Match 23.5%; Score 442; DB 2; Length 378;
Best Local Similarity 33.4%; Pred. No. 3.7e-25;
Matches 126; Conservative 59; Mismatches 140; Indels 52; Gaps 13;

Qy 9 LMWVAGTGGHVPGLAVAHHLMAQGHQV-----RWLGTADRMEDLVPKHGIEID 59
Db 6 VLLAAGGTGGHVPFGLAALAHKATKARGYQVHLVTDPSAERYAKGFPADSIHVVP SATI--- 62

Qy 60 FIRISLGRKGKIALIAAPLIRIFNARQARIMKAYKPDVVLGGVYSGPGGLAAM 116
Db 63 -----GSKNPISVVRSLWKLWGLRTARLVTKLKPVAVVGFVGITVFPPLAST 112

Search completed: June 7, 2004, 07:14:17
Job time : 24 secs

QY 117 SLGIPVVLHONGIAGLTNKLWARIATKVMQBPAGFPNAE-----VVGNPVRTDVL 168

DB 113 GLGVPSIIHQNAMVGRANK--ALAARVKAIAGGFLPPANGQYSEKTVATGNVPVPAVL 169

QY 169 A--LP-LPQORLAGREGPVRLVVGSGQCARILNTQMPQVAAKLGDS-----VIIHQSG 220

DB 170 AASEIPTYPSQ--TGE--TFQLVVFVFGSQGAQFFSSAVPAAICLMKDEQKRIVVTQOAR 225

QY 221 KGSQSQVEQAYABAGQPHKVTEFIDDMAAYAWADVVCERSGALTVEISEIAAGLPALFV 280

DB 226 PEDKDSVIASQYKLG-VKADVSPFFGDMASRIGEADLVISRSQASTVSELSVIGRPSILV 284

QY 281 PFQHK-DRQQVWNALPLEKAGAAKIIIEQPOLSDVDVANTLAGWSRE--TLLTWAERARAA 337

DB 285 PYPHALDHDQANNAALSAAGSASVVKQAEISPKLSLLSSALAEPRDLSATAAAKAT 344

QY 338 SIPDATERVANEVSRVA 354

DB 345 GKPHAADVLDLVEAIA 361

RESULT 20

C97611

hypothetical protein AGR_C3798 [imported] - Agrobacterium tumefaciens (strain C58, Cere

C:Species: Agrobacterium tumefaciens

C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002

C:Accession: C97611

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Aillinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A:Reference number: A97359; MUID:21608551; PMID:11743194

A:Accession: C97611

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-378 <KUR>

A:Cross-references: GB:AE007869; PIDN:AAK87844.1; PID:g15157228; GSPDB:GN00169

C:Genetics:

A:Gene: AGR_C3798

A:Map position: circular chromosome

C:Superfamily: murG protein

Query Match 23.5%, Score 442, DB 2; Length 378;

Best Local Similarity 33.4%; Pred.No.3.7e-25;

Matches 126; Conservative 59; Mismatches 140; Indels 52; Gaps 13;

QY 9 LVMWAGTGCHVFPGLAVAHLMAGQGV-----RWLGTADRMEADLVKPHGIEID 59

DB 6 VLLAAGTGCHVFPAEALATLTKRGVQVHLVTDSSRAERYAGKFPADIEHVFSATII-- 62

QY 60 FIRISLGRKGIKALIAAPLRIFNW---RQARIMKAYKIPDVVLGMGVGVSGPGGLAAW 116

DB 63 -----GSKNPISVWRSLWKLVLGRLTARLVTKLKPFVAVVGFGYTPVPPLLAST 112

QY 117 SLGIPVVLHONGIAGLTNKLWARIATKVMQBPAGFPNAE-----VVGNPVRTDVL 168

DB 113 GLGVPSIIHQNAMVGRANK--ALAARVKAIAGGFLPPANGQYSEKTVATGNVPVPAVL 169

QY 169 A--LP-LPQORLAGREGPVRLVVGSGQCARILNTQMPQVAAKLGDS-----VIIHQSG 220

DB 170 AASEIPTYPSQ--TGE--TFQLVVFVFGSQGAQFFSSAVPAAICLMKDEQKRIVVTQOAR 225

QY 221 KGSQSQVEQAYABAGQPHKVTEFIDDMAAYAWADVVCERSGALTVEISEIAAGLPALFV 280

DB 226 PEDKDSVIASQYKLG-VKADVSPFFGDMASRIGEADLVISRSQASTVSELSVIGRPSILV 284

QY 281 PFQHK-DRQQVWNALPLEKAGAAKIIIEQPOLSDVDVANTLAGWSRE--TLLTWAERARAA 337

DB 285 PYPHALDHDQANNAALSAAGSASVVKQAEISPKLSLLSSALAEPRDLSATAAAKAT 344

QY 338 SIPDATERVANEVSRVA 354

DB 345 GKPHAADVLDLVEAIA 361

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 07:01:04 ; Search time 17 Seconds
(without alignments)
1114.914 Million cell updates/sec

Title: US-09-829-275-1

Perfect score: 1881

Sequence: 1 MWGQGRKRLVMAGGTGGHV.....RVANEVSRVARALEHHHHH 364

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------|
| 1 | 1798 | 95.6 | 354 | 1 | MURG_ECOLI |
| 2 | 1793 | 95.3 | 355 | 1 | MURG_SHIFL |
| 3 | 1788 | 95.1 | 354 | 1 | MURG_ECO57 |
| 4 | 1786 | 94.9 | 354 | 1 | MURG_ECOL6 |
| 5 | 1682 | 89.4 | 354 | 1 | MURG_SALTY |
| 6 | 1672 | 88.9 | 354 | 1 | MURG_SALTI |
| 7 | 1471 | 78.2 | 356 | 1 | MURG_TERPE |
| 8 | 1431 | 76.1 | 355 | 1 | MURG_PHOOL |
| 9 | 1175 | 62.5 | 355 | 1 | MURG_VIBPA |
| 10 | 1166 | 62.0 | 355 | 1 | MURG_VIBVU |
| 11 | 1163 | 61.8 | 355 | 1 | MURG_VIBVY |
| 12 | 1154 | 61.4 | 354 | 1 | MURG_VIBCH |
| 13 | 1123 | 59.7 | 354 | 1 | MURG_PASNU |
| 14 | 1064.5 | 56.6 | 351 | 1 | MURG_HAEIN |
| 15 | 1021 | 54.3 | 355 | 1 | MURG_HAEUD |
| 16 | 945 | 50.2 | 354 | 1 | MURG_BUCAP |
| 17 | 939 | 49.9 | 354 | 1 | MURG_BUCAI |
| 18 | 900 | 47.8 | 360 | 1 | MURG_CANBF |
| 19 | 887 | 47.2 | 362 | 1 | MURG_SHRON |
| 20 | 869.5 | 46.2 | 365 | 1 | MURG_RALSO |
| 21 | 863.5 | 45.9 | 365 | 1 | MURG_SHEVI |
| 22 | 801 | 42.6 | 353 | 1 | MURG_WICBP |
| 23 | 798 | 42.4 | 359 | 1 | MURG_WIGBR |
| 24 | 798 | 42.4 | 364 | 1 | MURG_XYLFT |
| 25 | 794 | 42.2 | 357 | 1 | MURG_BORPA |
| 26 | 793 | 42.2 | 357 | 1 | MURG_BORBR |
| 27 | 791.5 | 42.1 | 426 | 1 | MURG_XANAC |
| 28 | 791 | 42.1 | 357 | 1 | MURG_BORPE |
| 29 | 787 | 41.8 | 358 | 1 | MURG_COXBU |
| 30 | 783.5 | 41.7 | 427 | 1 | MURG_XANCP |
| 31 | 780 | 41.5 | 367 | 1 | MURG_XYLF |
| 32 | 779.5 | 41.4 | 355 | 1 | MURG_NEIMB |
| 33 | 772.5 | 41.1 | 355 | 1 | MURG_NEIMA |

RESULT 1

| | | | | |
|-----|-------|------|---|------------|
| 34 | 764 | 40.6 | 1 | MURG_NITEU |
| 35 | 719.5 | 38.3 | 1 | MURG_PSEAE |
| 36 | 717 | 38.1 | 1 | MURG_PSESM |
| 37 | 683 | 36.3 | 1 | MURG_PSEPK |
| 38 | 510.5 | 27.1 | 1 | MURG_MICLE |
| 39 | 506 | 26.9 | 1 | MURG_MYCBO |
| 40 | 506 | 26.9 | 1 | MURG_MYCTU |
| 41 | 501.5 | 26.4 | 1 | MURG_BRUME |
| 42 | 496.5 | 26.7 | 1 | MURG_BRUSU |
| 43 | 487 | 25.9 | 1 | MURG_CORGL |
| 44 | 484 | 25.7 | 1 | MURG_CAUCR |
| 45 | 480.5 | 25.5 | 1 | MURG_COREF |
| 46 | 480 | 25.5 | 1 | MURG_RHIME |
| 47 | 476 | 25.3 | 1 | MURG_BACTN |
| 48 | 470 | 25.0 | 1 | MURG_BACHD |
| 49 | 470 | 25.0 | 1 | MURG_RHILO |
| 50 | 469 | 24.9 | 1 | MURG_PORGI |
| 51 | 462 | 24.6 | 1 | MUGI_BACCR |
| 52 | 461.5 | 24.5 | 1 | MURG_STRAW |
| 53 | 460 | 24.5 | 1 | MUGI_BACAA |
| 54 | 451 | 24.0 | 1 | MURG_SYNEL |
| 55 | 450.5 | 24.0 | 1 | MURG_STRCO |
| 56 | 447.5 | 23.8 | 1 | MURG_STRCU |
| 57 | 442 | 23.5 | 1 | MURG_AGRTS |
| 58 | 440 | 23.4 | 1 | MURG_GLOVI |
| 59 | 438 | 23.3 | 1 | MURG_THETN |
| 60 | 435 | 23.1 | 1 | MURG_LACLA |
| 61 | 433 | 23.0 | 1 | MURG_ZYMO |
| 62 | 431.5 | 22.9 | 1 | MURG_BIFLO |
| 63 | 425 | 22.6 | 1 | MURG_ANASP |
| 64 | 424 | 22.5 | 1 | MURG_LACPL |
| 65 | 422 | 22.4 | 1 | MURG_BACSU |
| 66 | 422 | 22.4 | 1 | MURG_LISMO |
| 67 | 419 | 22.3 | 1 | MURG_ENTHR |
| 68 | 418 | 22.2 | 1 | MURG_LISIN |
| 69 | 411.5 | 21.9 | 1 | MURG_ENTPA |
| 70 | 407 | 21.6 | 1 | MURG_CHLTE |
| 71 | 404 | 21.5 | 1 | MURG_BRAJA |
| 72 | 398.5 | 21.2 | 1 | MURG_SRYA3 |
| 73 | 340 | 18.1 | 1 | MURG_RICCN |
| 74 | 335 | 17.8 | 1 | MURG_OCEIH |
| 75 | 334.5 | 17.8 | 1 | MURG_AQAEH |
| 76 | 332.5 | 17.7 | 1 | MURG_CLOTE |
| 77 | 328.5 | 17.5 | 1 | MURG_CLOAB |
| 78 | 321.5 | 17.1 | 1 | MURG_THEMA |
| 79 | 319.5 | 17.0 | 1 | MURG_RICPR |
| 80 | 315.5 | 16.8 | 1 | MURG_TREPA |
| 81 | 313 | 16.6 | 1 | MURG_CHLTR |
| 82 | 306.5 | 16.3 | 1 | MUG2_BACCR |
| 83 | 304.5 | 16.2 | 1 | MURG_CLOPE |
| 84 | 304.5 | 16.2 | 1 | MURG_CLOPE |
| 85 | 303 | 16.1 | 1 | MURG_CHLMU |
| 86 | 299 | 15.9 | 1 | MURG_CHLCV |
| 87 | 284.5 | 15.1 | 1 | MURG_CHLPN |
| 88 | 277 | 14.7 | 1 | MURG_HELPY |
| 89 | 275 | 14.6 | 1 | MURG_HELPY |
| 90 | 273 | 14.5 | 1 | MURG_STAAM |
| 91 | 273 | 14.5 | 1 | MURG_STAAM |
| 92 | 272.5 | 14.5 | 1 | MURG_STRRG |
| 93 | 269.5 | 14.3 | 1 | MURG_STRPN |
| 94 | 263 | 14.0 | 1 | MURG_CAMJE |
| 95 | 261 | 13.9 | 1 | MURG_STAEP |
| 96 | 259.5 | 13.8 | 1 | MURG_BORBU |
| 97 | 250 | 13.3 | 1 | MURG_STRPY |
| 98 | 248 | 13.2 | 1 | MURG_STRP3 |
| 99 | 213.5 | 11.4 | 1 | MURG_STRMU |
| 100 | 150 | 8.0 | 1 | Y453_METAC |

ALIGNMENTS

MURG_ECOLI STANDARD; PRT; 354 AA.

AC P1743; 1990 (Rel. 15, Created)

DT 01-AUG-1990 (Rel. 30, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)

DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase

DE (SC 2.4.1.227) (Undecaprenyl-PP-MurNAC-pentapeptide-UDPglcNAC GlnNAC transferase).

DE MURG OR B0090.

GN Escherichia coli.

OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RP STRAIN=K12;

RX MEDLINE=90326550; PubMed=2197603;

RA Ikeda M., Wachi M., Jung H.K., Ishino F., Matsuhashi M.;

RT "Nucleotide sequence involving murG and murC in the mra gene cluster region of Escherichia coli.";

RL Nucleic Acids Res. 18:4014-4014(1990).

RN [2]

RP SEQUENCE FROM N.A.

RP STRAIN=K12;

RX MEDLINE=92334977; PubMed=1630901;

RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,

RT "Systematic sequencing of the Escherichia coli genome: analysis of the 0-2.4 min region.";

RL Nucleic Acids Res. 20:3305-3308(1992).

RN [3]

RP SEQUENCE FROM N.A.

RP STRAIN=K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1453-1474(1997).

RN [4]

RP SEQUENCE OF 8-354 FROM N.A.

RP STRAIN=K12;

RX MEDLINE=90251463; PubMed=2187180;

RA Mengin-Lecreulx D., Texier L., van Heijenoort J.;

RT "Nucleotide sequence of the cell-envelope murG gene of Escherichia coli.";

RL Nucleic Acids Res. 18:2810-2810(1990).

RN [5]

RP SEQUENCE OF 1-17, AND CHARACTERIZATION.

RX MEDLINE=91310568; PubMed=1649817;

RA Mengin-Lecreulx D., Texier L., Rousseau M., van Heijenoort J.;

RT "The murG gene of Escherichia coli codes for the UDP-N-acetylglucosamine: N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase involved in the membrane steps of peptidoglycan synthesis.";

RT J. Bacteriol. 173:4625-4636(1991).

RN [6]

RP SUBCELLULAR LOCATION.

RX MEDLINE=93194811; PubMed=8449890;

RA Bupp K., van Heijenoort J.;

RT "The final step of peptidoglycan subunit assembly in Escherichia coli occurs in the cytoplasm.";

RL J. Bacteriol. 175:1841-1843(1993).

RN [7]

RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).

RX MEDLINE=20348887; PubMed=10892798;

RA Ha S., Walker D., Shi Y., Walker S.;

RT "The 1.9 A crystal structure of Escherichia coli MurG, a membrane-associated glycosyltransferase involved in peptidoglycan biosynthesis.";

Protein Sci. 9:1045-1052(2000).

!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNAC subunit on undecaprenyl-pyrophosphoryl-MurNAC-pentapeptide (lipid intermediate I) to form undecaprenyl-pyrophosphoryl-MurNAC-(pentapeptide)GlcNAC (lipid intermediate II).

!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac(Oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP + GlcNAC-(1->4)-Mur2Ac(Oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol.

!- PATHWAY: Peptidoglycan biosynthesis; last step.

!- SUBCELLULAR LOCATION: Inner membrane-associated.

!- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG subfamily.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X52644; CAA36867.1; -

DR EMBL; X52540; CAA36776.1; -

DR EMBL; X55034; CAA38867.1; -

DR EMBL; D10483; BAB96658.1; -

DR EMBL; AE000118; AAC73201.1; -

PIR; J00544; BVECMG.

DR PDB; 1FOK; 27-JUL-00.

DR ECoGene; EG10623; murG.

DR HAMAP; MF_00033; -; 1.

DR InterPro; IPR007235; Glyco_tran_28_C.

DR InterPro; IPR004276; Glyco_trans_28.

DR InterPro; IPR006009; MurG_1.

PFam; PF04101; Glyco_tran_28_C; 1.

PFam; PF03033; Glyco_transf_28; 1.

TIGRFAMs; TIGR01133; murG; 1.

KW Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;

KW Inner membrane; Peptidoglycan synthesis; Complete proteome;

KW 3D-structure.

INIT_MET 0 0

STRAND 6 10

HELIX 15 29

TURN 30 32

TURN 34 39

TURN 41 42

TURN 44 44

TURN 45 48

TURN 49 52

TURN 53 53

TURN 55 58

TURN 65 66

TURN 69 73

TURN 74 74

TURN 76 93

TURN 97 100

TURN 104 105

TURN 106 115

TURN 116 117

TURN 120 124

TURN 131 136

TURN 137 139

TURN 142 145

TURN 148 149

TURN 155 156

TURN 163 166

TURN 167 167

TURN 171 175

TURN 176 177

TURN 182 187

TURN 189 191

TURN 194 207

TURN 208 210

```

FT STRAND 211 216
FT TURN 219 220
FT HELIX 222 231
FT TURN 222 231
FT TURN 232 233
FT TURN 235 236
FT STRAND 238 240
FT HELIX 246 252
FT STRAND 246 252
FT HELIX 255 258
FT TURN 262 271
FT TURN 272 272
FT STRAND 275 277
FT TURN 283 284
FT HELIX 286 296
FT TURN 297 298
FT STRAND 300 302
FT HELIX 305 307
FT TURN 310 318
FT TURN 319 319
FT HELIX 322 334
FT TURN 335 335
FT TURN 338 339
FT HELIX 340 352
FT TURN 353 354
SQ SEQUENCE 354 AA; 37683 MW; 0E3FAD945D769CID CRC64;

Query Match
Best Local Similarity 95.6%; Score 1798; DB 1; Length 354;
Matches 351; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 SQGQKRLVMAGGTGGHVFPGGLAVAHLMMAQGVQVRLGTADRMADLVPKHGIEDFI 62
DB 1 SQGQKRLVMAGGTGGHVFPGGLAVAHLMMAQGVQVRLGTADRMADLVPKHGIEDFI 60

QY 63 ISGLRGKIKALIAAPLRIFNARQARIMKAYKPDVVLGMGGVSGPGGLAAGSLGIP 122
DB 1 ISGLRGKIKALIAAPLRIFNARQARIMKAYKPDVVLGMGGVSGPGGLAAGSLGIP 120

QY 123 VLHEQNGIAGLTNKLWARIATKVMQAEPPAPNAEVVGNPVRTDVLALPLPQORLAGREG 182
DB 121 VLHEQNGIAGLTNKLWAKIATKVMQAEPPAPNAEVVGNPVRTDVLALPLPQORLAGREG 180

QY 183 PVRVLVVGSGQAGILNCTMPQVAAKLGDSTVIWHQSGKSGQSQVEQYAEAGQPQHKVT 242
DB 181 PVRVLVVGSGQAGILNCTMPQVAAKLGDSTVIWHQSGKSGQSQVEQYAEAGQPQHKVT 240

QY 243 EFIDDMAAYAWADVVCRSALTIVSEIAAGLPALFVFPQHKDRQOYWNALPLEKAGAA 302
DB 241 EFIDDMAAYAWADVVCRSALTIVSEIAAGLPALFVFPQHKDRQOYWNALPLEKAGAA 300

QY 303 KIIQPQLSDAVANTLAGSRETLLTWABRARAASIPDATERVANEVSVARA 356
DB 301 KIIQPQLSDAVANTLAGSRETLLTWABRARAASIPDATERVANEVSVARA 354

RESULT 2
MURG_SHIFL STANDARD; PRT; 355 AA.
AC Q8JNN4;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.22) (Undecaprenyl-PP-MurNac-pentapeptide-UDP-GlcNAc GlcNAc
DE transferase).
CS MURG OR SF0087 OR S0089.
GN Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;

```

```

RX MEDLINE=22272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding X., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157."
RL Nucleic Acids Res. 30:4432-4441(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T."
RL Infect. Immun. 71:2775-2786(2003).
CC -!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNAc
CC subunit on undecaprenyl-pyrophosphoryl-MurNac-pentapeptide (lipid
CC intermediate I) to form undecaprenyl-pyrophosphoryl-MurNac-
CC (pentapeptide)GlcNAc (lipid intermediate II) (By similarity).
CC -!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac(oyl-L-Ala-
CC gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP +
CC GlcNAc-(1->4)-Mur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
CC diphosphoundecaprenol.
CC -!- PATHWAY: Peptidoglycan biosynthesis; last step.
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG
CC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AB015046; AAN1752.1; -
DR EMBL; AB016978; AAP15633.1; -
DR HAMAP; MF_00033; -; 1.
DR InterPro; IPR004276; Glyco_trans_28.
DR InterPro; IPR007235; Glyco_trans_28_C.
DR InterPro; IPR006009; MurG_28.
DR Pfam; PF03033; Glyco_transf_28; 1.
DR Pfam; PF04101; Glyco_trans_28_C; 1.
DR TIGRFAMs; TIGR01133; murG; 1.
DR Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;
KW Inner membrane; Peptidoglycan synthesis; Complete proteome.
SQ SEQUENCE 355 AA; 37812 MW; 73407776C2B1504C CRC64;

Query Match
Best Local Similarity 95.3%; Score 1793; DB 1; Length 355;
Matches 350; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 MSGQKRLVMAGGTGGHVFPGGLAVAHLMMAQGVQVRLGTADRMADLVPKHGIEDFI 61
DB 1 MSGQKRLVMAGGTGGHVFPGGLAVAHLMMAQGVQVRLGTADRMADLVPKHGIEDFI 60

QY 62 RISGLRGKIKALIAAPLRIFNARQARIMKAYKPDVVLGMGGVSGPGGLAAGSLGIP 121
DB 61 RISGLRGKIKALIAAPLRIFNARQARIMKAYKPDVVLGMGGVSGPGGLAAGSLGIP 120

QY 122 VYLHEQNGIAGLTNKLWARIATKVMQAEPPAPNAEVVGNPVRTDVLALPLPQORLAGRE 181
DB 121 VYLHEQNGIAGLTNKLWAKIATKVMQAEPPAPNAEVVGNPVRTDVLALPLPQORLAGRE 180

QY 182 GPVRLVVGSGQAGILNCTMPQVAAKLGDSTVIWHQSGKSGQSQVEQYAEAGQPQHKV 241
DB 181 GPVRLVVGSGQAGILNCTMPQVAAKLGDSTVIWHQSGKSGQSQVEQYAEAGQPQHKV 240

```

QY 242 TEIDDMAAYAVADVVCVSGALTVSEIAAAGLPAALFVFFQHKDQOQYWNALPLEKAGA 301
 Db 241 TEIDDMAAYAVADVVCVSGALTVSEIAAAGLPAALFVFFQHKDQOQYWNALPLEKAGA 300
 QY 302 AKIEQPQLSVDVAVANTLACWSRETLITMAERARAASIPDATERVANEVSRVARA 356
 Db 301 AKIEQPQLSVDVAVANTLACWSRETLITMAERARAASIPDATERVANEVSRVARA 355

RESULT 3
 MURG ECO57 STANDARD; PRT; 354 AA.
 ID MURG ECO57
 AC Q8X9Y8;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE UDP-N-acetylglucosamine--N-acetylmuramyl- (pentapeptide)
 DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
 DE (EC 2.4.1.227) (Undecaprenyl-PP-MurNac-pentapeptide-UDP-GlcNAc GlcNAc transferase)
 GN MURG OR 20100 OR EGS0094.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=83334;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Hackett J., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
 RL Nature 403:529-533 (2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
 RA Kihara S., Shiba T., Hattori M., Shinagawa H.,
 RA "Complete genome sequence of enterohaemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12";
 RL DNA Res. 8:11-22 (2001).
 CC -!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNAc
 subunit on undecaprenyl-pyrophosphoryl-MurNac-pentapeptide (lipid
 intermediate I) to form undecaprenyl-pyrophosphoryl-MurNac-
 (pentapeptide)GlcNAc (lipid intermediate II) (By similarity).
 CC -!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac(oyl-L-Ala-
 gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP +
 GlcNAc-(1->4)-Mur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
 diphosphoundecaprenol.
 CC -!- PATHWAY: Peptidoglycan biosynthesis; last step.
 CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
 CC -!- SIMILARITY: Belongs to the glycosyltransferase family 28. Murg
 subfamily.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL; AE005185; AAG54394.1; --
 DR EMBL; AP002550; BAB33517.1; --
 PIR; P85491; P85491.

PIR; F90640; F90640.
 DR HAMAP; MF_00033; --; 1.
 DR InterPro; IPR007235; Glyco_tran_28_C.
 DR InterPro; IPR004276; Glyco_trans_28.
 DR InterPro; IPR006009; Murg.
 DR Pfam; PF04101; Glyco_tran_28_C; 1.
 DR Pfam; PF03033; Glyco_tran_28; 1.
 DR TIGRFAMs; TIGR01133; murg; 1.
 KW Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;
 KW Inner membrane; Peptidoglycan synthesis; Complete proteome.
 FT INIT MET 0 BY SIMILARITY.
 SQ SEQUENCE 354 AA; 37669 MW; 8C4617F0EBA10DE9 CRC64;

Query Match 95.1%; Score 1788; DB 1; Length 354;
 Best Local Similarity 98.6%; Pred. No. 9.7e-124;
 Matches 349; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 SQGKRLMVMAAGTGGVHVFGLAVAHLMAGQVWVRLGTADRMEDLVPKHGIEIDFIR 62
 Db 1 SAQKRLMVMAAGTGGVHVFGLAVAHLMAGQVWVRLGTADRMEDLVPKHGIEIDFIR 60
 QY 63 ISGLRGKGIKALIAAPLRFNARWQARIMKAYKPDVVVLGNGVYSGPGGLAANSLGIPV 122
 Db 61 ISGLRGKGIKALIAAPLRFNARWQARIMKAYKPDVVVLGNGVYSGPGGLAANSLGIPV 120
 QY 123 VLHEQNGIAGLTNKLWARIATKWAQEPGAPFNAEVVGNPVRTDVLALPLPQORLAGREG 182
 Db 121 VLHEQNGIAGLTNKLWARIATKWAQEPGAPFNAEVVGNPVRTDVLALPLPQORLAGREG 180
 QY 183 PVRVLVVGSGQCARILNQTMPQVAALGDSVTIWHQSGKSGSQSVQEQAYABAGQPHKVT 242
 Db 181 PVRVLVVGSGQCARILNQTMPQVAALGDSVTIWHQSGKSGSQSVQEQAYABAGQPHKVT 240
 QY 243 EPIDMAAYAWADVVCVSGALTVSEIAAAGLPAALFVFFQHKDQOQYWNALPLEKAGA 302
 Db 241 EPIDMAAYAWADVVCVSGALTVSEIAAAGLPAALFVFFQHKDQOQYWNALPLEKAGA 300

303 KIIEQPQLSVDVAVANTLACWSRETLITMAERARAASIPDATERVANEVSRVARA 356
 301 KIIEQPQLSVDVAVANTLACWSRETLITMAERARAASIPDATERVANEVSRVARA 354

RESULT 4
 MURG ECO57 STANDARD; PRT; 354 AA.
 ID MURG ECO57
 AC Q8FL64;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE UDP-N-acetylglucosamine--N-acetylmuramyl- (pentapeptide)
 DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
 DE (EC 2.4.1.227) (Undecaprenyl-PP-MurNac-pentapeptide-UDP-GlcNAc GlcNAc transferase)
 GN MURG OR C0108.
 OS Escherichia coli O6.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=217992;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O6:HI / CFT073 / ATCC 700928;
 RX MEDLINE=22388234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic Escherichia coli";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
 CC -!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNAc
 subunit on undecaprenyl-pyrophosphoryl-MurNac-pentapeptide (lipid
 intermediate I) to form undecaprenyl-pyrophosphoryl-MurNac-
 (pentapeptide)GlcNAc (lipid intermediate II) (By similarity).
 CC

```

CC CC -!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac (oyl-L-Ala-
CC CC gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP +
CC CC GlcNAc-(1->4)-Mur2Ac (oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
CC CC diphosphoundecaprenol.
CC CC -!- PATHWAY: Peptidoglycan biosynthesis; last step.
CC CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC CC -!- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG
CC CC subfamily.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC
CC CC EMBL; A016755; A016756.1; -.
CC CC HAMAP; MF_00033; -.
CC CC InterPro; IPR007235; Glyco_tran_28_C.
CC CC InterPro; IPR004276; Glyco_trans_28.
CC CC InterPro; IPR006009; MurG.
CC CC Pfam; PF04101; Glyco_tran_28_C; 1.
CC CC Pfam; PF03033; Glyco_transf_28; 1.
CC CC TIGRFAMs; TIGR01133; murG; 1.
CC CC Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;
CC CC Inner membrane; Peptidoglycan synthesis; Complete proteome.
CC CC INIT MET 0 BY SIMILARITY.
CC CC SEQUENCE 354 AA; 37645 MW; 0C74C03FDD769C1D CRC64;
CC CC
CC CC Query Match 94.9%; Score 1786; DB 1; Length 354;
CC CC Best Local Similarity 99.6%; Pred. No. 1.4e-123;
CC CC Matches 349; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
CC CC
CC CC QY 3 SQGKRLVMVAGGTGGHVFPGGLAVAHLMAGQVQVRLGTADRMEDLVPKHGIEDIFR 62
CC CC DB 1 SQGKRLVMVAGGTGGHVFPGGLAVAHLMAGQVQVRLGTADRMEDLVPKHGIEDIFR 60
CC CC
CC CC QY 63 ISGLRGKIKALIAAPLIRIFNANRQARIMKAYKPDVVLGMGGYVSGPGLAASLGIPV 122
CC CC DB 1 ISGLRGKIKALIAAPLIRIFNANRQARIMKAYKPDVVLGMGGYVSGPGLAASLGIPV 120
CC CC
CC CC QY 123 VLHONGIAGLTNKLARIATKWAEPGAPPAEAVVGNPVRTDVLALPLPQRLAGREG 182
CC CC DB 121 VLHONGIAGLTNKLAKIATVWQAPFGAPPAEAVVGNPVRTDVLALPLPQRLAGREG 180
CC CC
CC CC QY 183 PVRVLVVGSGQARILNQTMTPOVAALGDSVITWHQSGKGSQSQVQAYAEAGQPQHKVT 242
CC CC DB 181 PVRVLVVGSGQARILNQTMTPOVAALGDSVITWHQSGKGSQSQVQAYAEAGQPQHKVT 240
CC CC
CC CC QY 243 EFIDDMAAYAWADVVCRSGLTVSEIAAAGLPALFVFPQHKDROQYWNALPLEKAGAA 302
CC CC DB 241 EFIDDMAAYAWADVVCRSGLTVSEIAAAGLPALFVFPQHKDROQYWNALPLEKAGAA 300
CC CC
CC CC RESULT 5
CC CC MURG SALTY STANDARD; PRT; 354 AA.
CC CC AC Q82RU3;
CC CC DT 28-FEB-2003 (Rel. 41, Created)
CC CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
CC CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
CC CC DE UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)
CC CC DE pyrophosphoryl-undecaprenol-N-acetylglucosamine transferase
CC CC DE (EC 2.4.1.22) (Undecaprenyl-pp-MurNAC-pentapeptide-UDP-GlcNAc GlcNAc
CC CC transferase).
CC CC GN MURG OR STM0128
CC CC OS Salmonella typhimurium.
CC CC OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

```

```

OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856 (2001).
CC -!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNAc
CC subunit on undecaprenyl-pyrophosphoryl-MurNAC-pentapeptide (lipid
CC intermediate I) to form undecaprenyl-pyrophosphoryl-MurNAC-
CC (pentapeptide/GlcNAc (lipid intermediate II) (By similarity).
CC -!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac (oyl-L-Ala-
CC gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP +
CC GlcNAc-(1->4)-Mur2Ac (oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
CC diphosphoundecaprenol.
CC -!- PATHWAY: Peptidoglycan biosynthesis; last step.
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG
CC subfamily.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC
CC CC EMBL; A008699; AAL19092.1; -.
CC CC HAMAP; MF_00033; -.
CC CC InterPro; IPR007235; Glyco_tran_28_C.
CC CC InterPro; IPR004276; Glyco_trans_28.
CC CC InterPro; IPR006009; MurG.
CC CC Pfam; PF04101; Glyco_tran_28_C; 1.
CC CC Pfam; PF03033; Glyco_transf_28; 1.
CC CC TIGRFAMs; TIGR01133; murG; 1.
CC CC Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;
CC CC Inner membrane; Peptidoglycan synthesis; Complete proteome.
CC CC INIT MET 0 BY SIMILARITY.
CC CC SEQUENCE 354 AA; 37731 MW; 0E3850B6C6CDFD2E CRC64;
CC CC
CC CC Query Match 89.4%; Score 1682; DB 1; Length 354;
CC CC Best Local Similarity 91.2%; Pred. No. 5.5e-116;
CC CC Matches 322; Conservative 18; Mismatches 13; Indels 0; Gaps 0;
CC CC
CC CC QY 3 SQGKRLVMVAGGTGGHVFPGGLAVAHLMAGQVQVRLGTADRMEDLVPKHGIEDIFR 62
CC CC DB 1 SQGKRLVMVAGGTGGHVFPGGLAVAHLMAGQVQVRLGTADRMEDLVPKHGIEDIFR 60
CC CC
CC CC QY 63 ISGLRGKIKALIAAPLIRIFNANRQARIMKAYKPDVVLGMGGYVSGPGLAASLGIPV 122
CC CC DB 61 ISGLRGKIKALIAAPLIRIFNANRQARIMKAYKPDVVLGMGGYVSGPGLAASLGIPV 120
CC CC
CC CC QY 123 VLHONGIAGLTNKLARIATKWAEPGAPPAEAVVGNPVRTDVLALPLPQRLAGREG 182
CC CC DB 121 VLHONGIAGLTNKLAKIATVWQAPFGAPPAEAVVGNPVRTDVLALPLPQRLAGREG 180
CC CC
CC CC QY 183 PVRVLVVGSGQARILNQTMTPOVAALGDSVITWHQSGKGSQSQVQAYAEAGQPQHKVT 242
CC CC DB 181 PVRVLVVGSGQARILNQTMTPOVAALGDSVITWHQSGKGSQSQVQAYAEAGQPQHKVT 240
CC CC
CC CC QY 243 EFIDDMAAYAWADVVCRSGLTVSEIAAAGLPALFVFPQHKDROQYWNALPLEKAGAA 302
CC CC DB 241 EFIDDMAAYAWADVVCRSGLTVSEIAAAGLPALFVFPQHKDROQYWNALPLEKAGAA 300

```

QY 303 KITEPOLSDVAVNTLAGSRETLTMAERAAASIPDATERVANEVSRRVAR 355
 Db 301 KIFEQCFTEVAVADTLAGSREALTMAERAAVSPDATERVASEVSRRVAR 353

RESULT 6

MURG_SALTI STANDARD; PRT; 354 AA.
 AC Q82939; 2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)
 DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
 DE (EC 2.4.1.227) [Undecaprenyl-PP-MurNac-pentapeptide-UDP-GlcNAc GlcNAc transferase].
 GN MURG OR STY0148 OR T0132.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_TaxID=601;
 [1]
 SEQUENCE FROM N.A.
 RP STRAIN=CT18;
 RC MEDLINE=21534947; PubMed=11677608;
 RX Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Rarrar J.,
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagsis K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Farry C.,
 RA Quail M.A., Rutherford K., Sammonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrall B.G.;
 RA "Complete genome sequence of a multiple drug resistant Salmonella
 enterica serovar Typhi CT18";
 RT Nature 413:848-852(2001).
 RL [2]
 SEQUENCE FROM N.A.
 RP STRAIN=Ty2 / ATCC 700931;
 RC MEDLINE=22531367; PubMed=12644504;
 RX Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
 RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
 RA "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
 RT and CT18";
 RL J. Bacteriol. 185:2330-2337(2003).
 CC -|- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNAc
 subunit on undecaprenyl-pyrophosphoryl-MurNac-pentapeptide (lipid
 intermediate I) to form undecaprenyl-pyrophosphoryl-MurNac-
 (pentapeptide)GlcNAc (lipid intermediate II) (By similarity).
 CC -|- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac(oyl-L-Ala-
 gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP +
 GlcNAc-(1->4)-Mur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
 diphosphoundecaprenol.
 CC -|- PATHWAY: Peptidoglycan biosynthesis; last step.
 CC -|- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
 CC -|- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG
 subfamily.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; AL627265; CAD01295.1; -;
 EMBL; AE016834; AAC67864.1; -;
 HMAP; MF 00033; -; 1.
 InterPro; IPR007235; Glyco_tran.28.C.
 InterPro; IPR004276; Glyco_tran.28.
 InterPro; IPR006009; MurG.
 Pfam; PF04101; Glyco_tran.28.C; 1.

DR PF03033; Glyco transf.28; 1.
 DR TIGRfam; TIGR01133; murG; 1
 KW Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;
 KW inner membrane; Peptidoglycan synthesis; Complete proteome.
 FT INIT MET 0 BY SIMILARITY.
 SQ SEQUENCE 354 AA; 37689 MW; 9BE582B5E69BBD72 CRC64;

Query Match 88.9%; Score 1672; DB 1; Length 354;
 Best Local Similarity 90.7%; Pred. No. 3e-115;

Matches 320; Conservative 19; Mismatches 14; Indels 0; Gaps 0;
 QY 3 SQGKRLVMAGTGGHVFPGLAVALHLMACGQVQRWLTADNRMEADLVKPHGIEIDFIR 62
 Db 1 SQGPKRLVMAGTGGHVFPGLAVALHLMACGQVQRWLTADNRMEADLVKPHGIDIDFIR 60
 QY 63 ISGLRGKIGKALIAAPLRIFNANRQARAIMKAYKPDVVLGMMGGVSGPGGLAWSLGPV 122
 Db 61 ISGLRGKGVKALLAAPLRIFNANRQARAIMKRFKPDVVLGMMGGVSGPGGLAWSLGPV 120
 QY 123 VLHEQNGIAGLTNKLWARIATKVMQAPFPNNAEVVGNFVRTDVLALPLPQORLAGREG 182
 Db 121 VLHEQNGIAGLTNQWLAKIATTVNQAFPNAEVVGNFVRTDVLALPLPQVRLAGRDG 180
 QY 183 PVRVLVVGSSQCARILNQTPQVAAKLGDSVIWHQSGKSQSQSVQAYAEAGOPQHKT 242
 Db 181 PIRVLVVGSSQCARVLNQTPQVAARLGDTVTIWHQSGKAQLTVQAYAGAGOPQHKT 240
 QY 243 EFIDMAAAAYAWADVVCRSGLATVSEIAAAGLPALFPFQHKDRQOYWNALPLEKAGAA 302
 Db 241 EFIDGMAAAAYAWADVVCRSGLATVSEIAAAGLPALFPFQHKDRQOYWNALPLENAGAA 300
 QY 303 KIIPQOLSVDVANTLAGSRETLTMAERAAASIPDATERVANEVSRRVAR 355
 Db 301 KIFEQCFTEVAVADTLAGSREALTMAERAAVSPDATERVASEVSRRVAR 353

RESULT 7

MURG_YERPE STANDARD; PRT; 356 AA.
 AC Q82IE9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)
 DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
 DE (EC 2.4.1.227) [Undecaprenyl-PP-MurNac-pentapeptide-UDP-GlcNAc GlcNAc transferase].
 DE MURG OR YPO0555 OR Y3626.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Yersinia.
 OX NCBI_TaxID=632;
 [1]
 SEQUENCE FROM N.A.
 RP STRAIN=CO-92 / Biovar Orientalis;
 RC MEDLINE=21470413; PubMed=11586360;
 RX Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
 RA Sebahia M., Sebahia M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Feltwell T., Hamlin N., Holroyd S., Jagsis K., Karlyshev A.V.,
 RA Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.,
 RA Sammonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
 RT "Genome sequence of Yersinia pestis, the causative agent of plague";
 RL Nature 413:523-527(2001).
 [2]
 SEQUENCE FROM N.A.
 RP STRAIN=KIM5 / Biovar Mediaevalis;
 RC MEDLINE=22137863; PubMed=12142430;
 RX Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
 RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
 RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
 RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,


```

QY 302 AKIEPOLSDVAVANTLWAGSRETLTMAERARASIPDATERVA 347
DB 301 AKILEQOFTVDVIELTQWQPQLLEMAKARSAAIVDATEQVS 346

RESULT 9
MURG_VIBVU
ID MURG_VIBVU STANDARD; PRT; 355 AA.
AC Q87SG4;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.227) (Undecaprenyl-PP-MurNAC-pentapeptide-UDP-GlcNAc
DE transferase).
DE MURG OR V70460.
GN Vibrio parahaemolyticus.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIWD 2210633 / Serotype O3:K6;
RX MEDLINE=22508454; PubMed=12620739;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Najima M., Nakano M., Yoshishita A., Kubota Y., Kimura S.,
RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RA "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT distinct from that of V. cholerae";
RL Lancet 361:743-749(2003).
RC -!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNAc
subunit on undecaprenyl-pyrophosphoryl-MurNAC-pentapeptide (lipid
intermediate I) to form undecaprenyl-pyrophosphoryl-MurNAC-
(pentapeptide)GlcNAc (lipid intermediate II) (By similarity).
CC -!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac(oyl-L-Ala-
gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP +
GlcNAc-(1->4)-Mur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
diphosphoundecaprenol.
CC -!- PATHWAY: Peptidoglycan biosynthesis; last step.
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG
subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; AP005074; BAC58723.1; -.
CC HAMAP; MF_00033; -; 1.
CC InterPro; IPR007235; Glyco_tran_28.C.
CC DR InterPro; IPR004276; Glyco_trans_28.
CC DR Pfam; PF04101; Glyco_tran_28.C; 1.
CC DR Pfam; PF03033; Glyco_transf_28; 1.
CC DR Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;
KW Inner membrane; Peptidoglycan synthesis; Complete proteome.
SQ SEQUENCE 355 AA; 38250 MW; D938001FC452BDF2 CRC64;

Query Match
Best local similarity 62.5%; Score 1175; DB 1; Length 355;
Matches 226; Conservative 48; Mismatches 72; Indels 0; Gaps 0;

QY 5 QGKRLMNVAGTGGHVFPGLAVAHLMNAQGVWLGTADRMEDADLPVKIGIIFIRIS 64
DB 3 QNKRLMNVAGTGGHVFPGLAVAKLOEQGQWIEIWLGTADRMEDADLPVKIGIIFIKV 62
QY 65 GLRGKGIKALTAAPLRIFNWQRARIMKAYKPDVLGMGYVSGPGGLAAWSLGPVVL 124

```

Query Match 62.0%; Score 1166; DB 1; Length 355;
Best Local Similarity 65.9%; Pred. No. 3.2e-78;
Matches 228; Conservative 47; Mismatches 71; Indels 0; Gaps 0;

OY 5 QGKRLMWAGGTGGHVFPGGLAVAHHLMAQGMQVWLGTADRMADLVPKHGIEIDFIRIS 64
DB 3 KNKRLMWAGGTGGHVFPGGLAVAKKLOQGWGWEIRWLTADRMADLVPKHGIDIDIKVK 62

OY 65 GLRGKIKALIAAPLIRFNWQRAIMKAYKPDVVLGMGGVSGPGGLAWSLGPVVL 124
DB 63 GLRGOGIKRLVLAPFQILNAIFQAKHAKRQWPDVVLGMGGVSGPGGLAWSLGPVVL 122

OY 125 HEQNGIAGLTNKLWARIATKVMQAEFGAPFNAEVVGNVPTDVLALPLPQORLAGREGPV 184
DB 123 HEQNAVAGLTNHLWAKIAKVFQAFPGAFKQAPVVGPNVREDVVALPDMQRMQDREGAV 182

OY 185 RVLVVGSGQARILNQTMPQVAAKLGDVSVIIHQSGKSGQSQSVQAYAEAGQPHKVTFF 244
DB 183 RILVVGSGQARILNQTMPQVAAKLGDVSVIIHQSGKSGQSQSVQAYAEAGQPHKVTFF 242

OY 245 IDDMAAYAWADVVCRSGLTIVSEIAAGLPALFVFPQHKDQOYVWALPLEKAGAARI 304
DB 243 IDVAAQYAWADLVVCRSGALTIVSEVSAAGVGAIFIPFMHQRQOALNADHLVACGAALM 302

OY 305 IEQPOLSDVAVANTLAGRSRETLTMAERARAASIPDATERVANEV 350
DB 303 IEQPOLTVDKLAGIQKLGRTLLSMALHARAAQNNADQVADAI 348

RESULT 11
MURG_VIBVY STANDARD; PRT; 355 AA.

AC Q7NNVL;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.227) (Undecaprenyl-PP-MurNac-pentapeptide-UDP-GlcNAc
DE transferase).
GN MURG OR VV0614.
OS Vibrio vulnificus (strain YJ016).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrrio.
OX NCBI_TaxID=196600;
RN [1]
RP Chen C.Y., Wu X.M., Chang Y.C., Chang C.H., Tsai H.C., Liao T.L.,
RA Liu Y.M., Chen H.J., Shen A.B., Li J.C., Su T.L., Shao C.P., Lee C.T.,
RA Hor L.I., Tsai S.F.;
RT "Comparative genome analysis of Vibrio vulnificus, a marine
pathogen."
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNAc
CC subunit on undecaprenyl-pyrophosphoryl-MurNac-pentapeptide (lipid
CC intermediate I) to form undecaprenyl-pyrophosphoryl-MurNac-
CC (pentapeptide)GlcNAc (lipid intermediate II) (By similarity).
CC -!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac(oyl-L-Ala-
CC gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP +
CC GlcNAc-(1->4)-Mur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
CC diphosphoundecaprenol.
CC -!- PATHWAY: Peptidoglycan biosynthesis; last step.
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyltransferase family 28. Murg
CC subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AP05332; BAC93378.1; ALT_INIT.
DR HAMAP; MF_00033; -; 1.
KW Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;
KW Inner membrane; Peptidoglycan synthesis; Complete proteome.
SQ SEQUENCE 355 AA; 38170 MW; 1CED565A38A9D120 CRC64;

Query Match 61.8%; Score 1163; DB 1; Length 355;
Best Local Similarity 65.9%; Pred. No. 5.3e-78;
Matches 228; Conservative 46; Mismatches 72; Indels 0; Gaps 0;

OY 5 QGKRLMWAGGTGGHVFPGGLAVAHHLMAQGMQVWLGTADRMADLVPKHGIEIDFIRIS 64
DB 3 KNKRLMWAGGTGGHVFPGGLAVAKKLOQGWGWEIRWLTADRMADLVPKHGIDIDIKVK 62

OY 65 GLRGKIKALIAAPLIRFNWQRAIMKAYKPDVVLGMGGVSGPGGLAWSLGPVVL 124
DB 63 GLRGOGIKRLVLAPFQILNAIFQAKHAKRQWPDVVLGMGGVSGPGGLAWSLGPVVL 122

OY 125 HEQNGIAGLTNKLWARIATKVMQAEFGAPFNAEVVGNVPTDVLALPLPQORLAGREGPV 184
DB 123 HEQNAVAGLTNHLWAKIAKVFQAFPGAFKQAPVVGPNVREDVVALPDMQRMQDREGAV 182

OY 185 RVLVVGSGQARILNQTMPQVAAKLGDVSVIIHQSGKSGQSQSVQAYAEAGQPHKVTFF 244
DB 183 RILVVGSGQARILNQTMPQVAAKLGDVSVIIHQSGKSGQSQSVQAYAEAGQPHKVTFF 242

OY 245 IDDMAAYAWADVVCRSGLTIVSEIAAGLPALFVFPQHKDQOYVWALPLEKAGAARI 304
DB 243 IDVAAQYAWADLVVCRSGALTIVSEVSAAGVGAIFIPFMHQRQOALNADHLVACGAALM 302

OY 305 IEQPOLSDVAVANTLAGRSRETLTMAERARAASIPDATERVANEV 350
DB 303 IEQPOLTVDKLAGIQKLGRTLLSMALHARAAQNNADQVADAI 348

RESULT 12
MURG_VIBCH STANDARD; PRT; 354 AA.

AC Q9KPG7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.227) (Undecaprenyl-PP-MurNac-pentapeptide-UDP-GlcNAc
DE transferase).
GN MURG OR VC2401.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
CC STRAIN=El Tor N16961 / Serotype O1;
CC MEDLINE=20406833; PubMed=10952301;
CC Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
CC Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
CC Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
CC McDaniel L., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
CC McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
CC Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
CC Fraser C.M.;
CC "DNA sequence of both chromosomes of the cholera pathogen Vibrio
CC cholerae."
CC Nature 406:477-483 (2000).
CC -!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNAc
CC subunit on undecaprenyl-pyrophosphoryl-MurNac-pentapeptide (lipid
CC intermediate I) to form undecaprenyl-pyrophosphoryl-MurNac-
CC (pentapeptide)GlcNAc (lipid intermediate II) (By similarity).
CC -!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac(oyl-L-Ala-
CC gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP +
CC GlcNAc-(1->4)-Mur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
CC diphosphoundecaprenol.

```
CC diposphoundecaprenol.
CC -!- PATHWAY: Peptidoglycan biosynthesis; last step.
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated. (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG
CC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE004310; AA955544.1; -
CC PIR: E82081; E82081.
CC HSSP: P17443; 1FOK.
CC TIGR: VC2401; -.
CC HAMAP: MF_00033; -.
CC InterPro: IPR007235; Glyco_tran_28_C.
CC InterPro: IPR004276; Glyco_trans_28.
CC InterPro: IPR006009; MurG.
CC Pfam: PF04101; Glyco_tran_28_C; 1.
CC Pfam: PF03033; Glyco_transf_28; 1.
CC TIGRFAMs: TIGR01133; murG; 1.
CC Transferase; Glycosyltransferase; Cell division; Cell wall;
CC Inner membrane; Membrane; Peptidoglycan synthesis; Complete proteome.
CC SEQUENCE 354 AA; 37983 MW; 7B5C773B24476410 CRC64;
CC -----
Query Match 61.4%; Score 1154; DB 1; Length 354;
Best Local Similarity 64.2%; Pred. No. 2.4e-77;
Matches 224; Conservative 48; Mismatches 77; Indels 0; Gaps 0;
QY 2 MSGGKRLVWVAGTGGHVFPGLAVAHLMQAQGVQVRLGTADRMEADLVPKHGIEIDFI 61
DB 1 MNNKNNKLLWVAGTGGHVFPGLAVAKQLQQGQVIRLWLTADMEAEALVPKHGIEIDFI 60
QY 62 RISGLRGKIGKALIAAPLRIFNAMRQARAIMKAYKPDVIVGMGYVGGPGGLAANSLGIP 121
DB 61 QVKGILRGQGLMRLKAPFQIVNAILQARRHLITVQPDVILGMGYVGGPGGLAANSLGIP 120
QY 122 VVLHEQNGIAGLTNKKWLARIATKVMQAEPPGAFFNAEVVGNPVRTDVLALPLPQORLAGRE 181
DB 121 VVLHEQNAVALTNQWLAKIARRVQLPFGAPFADASVGNPVRTDVLALPLPQORLAGRE 180
QY 182 GPRVRLVVGSGQGARILNQTMPQVAALGDSVLIWHQSGKSGSQSVQEAAYAEAGQPQHKV 241
DB 181 GAIRILVVGSGQGARILNQTILPAVMAALGEGYEIRHQAGKNSQODVAEAYAAAGVESAQV 240
QY 242 TEFTDDMAAYANADVVVCRSGALTIVSEIAAAGLPALFVFPQHKDQOQYWNALPLEKAGA 301
DB 241 TEFTDDVADAYAWADLLICRSGALTIVSEVSAAGVGAIFIPFMKDRQOALNADHLVACGA 300
QY 302 AKIIEQPQLSVDAVANTLAGWSRETLTMAERAAASIPDATERVANEV 350
DB 301 AKMIEQPQLSVKELTQWVRELDRAQLLSMAQKAPQAALDADKVAQAI 349
RESULT 13
MURG_PASMU STANDARD; PRT; 354 AA.
AC P57817;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmutarimyl-(pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.227) (Undecaprenyl-PP-MurNac-pentapeptide-UDPglcNac GlcNac
DE transferase).
CN MURG OR PNO142.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
```

```
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Pm70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNac
CC subunit on undecaprenyl-pyrophosphoryl-MurNac-pentapeptide (lipid
CC intermediate 1) to form undecaprenyl-pyrophosphoryl-MurNac-
CC (pentapeptide)GlcNac (lipid intermediate 1) (By similarity).
CC -!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac(oyl-L-Ala-
CC gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP +
CC GlcNac- (1->4)-Mur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
CC diphosphoundecaprenol.
CC -!- PATHWAY: Peptidoglycan biosynthesis; last step.
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG
CC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE006049; AAK02226.1; -
CC HSSP: P17443; 1FOK.
CC HAMAP: MF_00033; -.
CC InterPro: IPR007235; Glyco_tran_28_C.
CC InterPro: IPR004276; Glyco_trans_28.
CC InterPro: IPR006009; MurG.
CC Pfam: PF04101; Glyco_tran_28_C; 1.
CC Pfam: PF03033; Glyco_transf_28; 1.
CC TIGRFAMs: TIGR01133; murG; 1.
CC Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;
CC Inner membrane; Peptidoglycan synthesis; Complete proteome.
CC SEQUENCE 354 AA; 38138 MW; D6F7EB5DF69A46C CRC64;
CC -----
Query Match 59.7%; Score 1123; DB 1; Length 354;
Best Local Similarity 63.1%; Pred. No. 4.4e-75;
Matches 222; Conservative 46; Mismatches 76; Indels 8; Gaps 2;
QY 2 MSGGKRLVWVAGTGGHVFPGLAVAHLMQAQGVQVRLGTADRMEADLVPKHGIEIDFI 61
DB 1 NSEQKRLVWVAGTGGHVFPGLAVAHLMQAQGVQVRLGTADRMEADLVPKHGIEIDFI 60
QY 62 RISGLRGKIGKALIAAPLRIFNAMRQARAIMKAYKPDVIVGMGYVGGPGGLAANSLGIP 121
DB 61 QISGLRGKIGKALIAAPLRIFNAMRQARAIMKAYKPDVIVGMGYVGGPGGLAANSLGIP 120
QY 122 VVLHEQNGIAGLTNKKWLARIATKVMQAEPPGAFFNAEVVGNPVRTDVLALPLPQORLAGRE 181
DB 121 VVLHEQNAVALTNQWLAKIARRVQLPFGAPFADASVGNPVRTDVLALPLPQORLAGRE 180
QY 182 GPRVRLVVGSGQGARILNQTMPQVAALGDSVLIWHQSGKSGSQSVQEAAYAEAGQPQHKV 239
DB 181 GKRLVRLVVGSGQGARILNQTILPAVMAALGEGYEIRHQAGKNSQODVAEAYAAAGVESAQV 235
QY 240 -KVTEFIDDDMAAYANADVVVCRSGALTIVSEIAAAGLPALFVFPQHKDQOQYWNALPLEK 298
DB 236 VTEFIDDDMAAYANADVVVCRSGALTIVSEIAAAGLPALFVFPQHKDQOQYWNALPLEK 295
QY 299 AGAIIQPOLSVDAVANTLAGWSRETLTMAERAAASIPDATERVANEV 350
DB 296 AGAIIQPOLSVDAVANTLAGWSRETLTMAERAAASIPDATERVANEV 347
RESULT 14
MURG_HAEIN
```

```
ID AC MURG_HAEIN STANDARD; PRT; 351 AA.
DT P45065;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.227) (Undecaprenyl-PP-MurNac-pentapeptide-UDP-GlcNac GlcNac
DE transferase).
DE MURG OR H1138.
GN Haemophilus influenzae.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
EX MEDLINE=95350630; PubMed=7542600;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512(1995).
CC -!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNac
CC subunit on undecaprenyl-pyrophosphoryl-MurNac-pentapeptide (lipid
CC intermediate I) to form undecaprenyl-pyrophosphoryl-MurNac-
CC (pentapeptide)GlcNac (lipid intermediate II) (By similarity).
CC -!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac(oyl-L-Ala-
CC gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP +
CC GlcNac-(1->4)-Mur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
CC diphosphoundecaprenol.
CC -!- PATHWAY: Peptidoglycan biosynthesis; last step.
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated. (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32793; AAC22793.1; -
CC PIR; D64185; D64185.
CC HSSP; P17443; 1FOK.
CC TIGR; H11138; -.
CC HAMAP; MF 00033; -.
CC InterPro; IPR007235; Glyco_tran_28_C.
CC InterPro; IPR004276; Glyco_tran_28.
CC InterPro; IPR006009; MurG.
CC Pfam; PF04101; Glyco_tran_28_C; 1.
CC Pfam; PF03033; Glyco_transf_28; 1.
CC TIGRFAMs; TIGR01133; murG; 1.
CC TrEMBL; F03033; Glyco_transf_28; 1.
KW Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;
KW Inner membrane; Peptidoglycan synthesis; Complete proteome.
SQ SEQUENCE 351 AA; 38307 MW; 41067F379BB30B27 CRC64;
Query Match 56.6%; Score 1064.5; DB 1; Length 351;
Best Local Similarity 61.5%; Pred. No. 8.4e-71;
Matches 214; Conservative 44; Mismatches 89; Indels 1; Gaps 1;
QY 5 OCKRLVWAGGTGGHVPGLVAHLMAGQWQVRLGTADRMEDLVKPKHIEIDFIRIS 64
DB 2 KKKLLVWAGGTGGHVPFAIAVAQTLQKQWDICLSTKDRMEALVKPKYGIPIRFQIS 61
```

```
QY 65 GLRGKGIKALIAAPLRIFNAMQARAIMKAYKPDVVLGNGGVSGPGGLAAMSLGIPVL 124
DB 62 GLRGKGIKALLNAPFAIRAVLQAKKIQEERPDVAVLGNGGVSGPAGVAAKLCGPVIL 121
QY 125 HQNGIAGTNTKWLARIATKWQAEFGAPNPAFVGNVPTDVLALPLPQORLAGSGPV 184
DB 122 HQCNAIAGTNTKLLGKIATCVLQAFPTAPPFAEVGNVPREDLFEMPAPDIFRSREK 181
QY 185 RVLVVGSGQCARILNQTMPQVAAKLGDSVIWHQSGKSGSQSQVEGAYAEAGQFQHKVTEF 244
DB 182 RVLVVGSGQCARVNLHTLPKVVAQLADKLEPFGVKGAVEVSQLYGE-NLEQVKITEF 240
QY 245 IDMAAAAYAMADVVCRSALTIVSIIAAGLPALFVPOHKDRQOYVWALPLEKGAARI 304
DB 241 IDNMAEYAWADVVCRSALTIVCEIAAVGAAGAAIFVFPQHKDRQOYVWALPLEKGAARI 300
QY 305 ISOPOLSDVADVANTLAGWSRETLTMAERARAASIPDATERVANEVSR 352
DB 301 IEQADLTPEILVNYLKNLTRENLQVALKATMSMPNAAQORVAEVIKQ 348
RESULT 15
MURG_HAEIDU STANDARD; PRT; 355 AA.
AC Q7U336;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.227) (Undecaprenyl-PP-MurNac-pentapeptide-UDP-GlcNac GlcNac
DE transferase).
DE MURG OR HD0824.
GN Haemophilus ducreyi.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3500HP / ATCC 700724;
RA Munson R.S. Jr., Ray W.C., Mahairas G., Sabo P., Mungur R.,
RA Johnson L., Nguyen D., Wang J., Forst C., Hood L.;
RT "The complete genome sequence of Haemophilus ducreyi."
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNac
CC subunit on undecaprenyl-pyrophosphoryl-MurNac-pentapeptide (lipid
CC intermediate I) to form undecaprenyl-pyrophosphoryl-MurNac-
CC (pentapeptide)GlcNac (lipid intermediate II) (By similarity).
CC -!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac(oyl-L-Ala-
CC gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP +
CC GlcNac-(1->4)-Mur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
CC diphosphoundecaprenol.
CC -!- PATHWAY: Peptidoglycan biosynthesis; last step.
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated. (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE017153; AAP95721.1; -
CC HAMAP; MF 00033; -.
CC InterPro; IPR004276; Glyco_tran_28.
CC InterPro; IPR007235; Glyco_tran_28_C.
CC InterPro; IPR006009; MurG.
CC Pfam; PF04101; Glyco_tran_28_C; 1.
CC Pfam; PF03033; Glyco_transf_28; 1.
CC TIGRFAMs; TIGR01133; murG; 1.
KW Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;
KW Inner membrane; Peptidoglycan synthesis; Complete proteome.
SQ SEQUENCE 351 AA; 38307 MW; 41067F379BB30B27 CRC64;
```

DR TIGRFAMS; TIGR01133; murG; 1.
KW Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;
KW Inner membrane; Peptidoglycan synthesis; Complete proteome.
SQ SEQUENCE 355 AA; 38478 MW; 5B74CBE2BFB6574 CRC64;

Query Match 54.3%; Score 1021; DB 1; Length 355;
Best Local Similarity 57.5%; Pred. No. 1.3e-67;
Matches 204; Conservative 51; Mismatches 92; Indels 8; Gaps 3;

QY 7 KRLVMAGGTHGVFPGGLAVAHLMAGQWVRVLGTADRMADLVPHKGIEIDFIRISGL 66
DB 3 KLLLMAGGTHGVFPAIAVAQELQKQWICWLTGKDRNEALVQVNIPIEFIOISGL 62
QY 67 RKGKIKALIAAPLIFINAWQARAIMKAYKDPVVLGMGVSGPGGLAWSIGIPVLHE 126
DB 63 KKGVLALIKAPPTILKAVLQALNIIKKYRPAVLGMGVSGPGGIAARLNCNPVLHE 122
QY 127 QNGTAGLTKNLARIATKVMQAPGAPNAEVGNVPTDVLALPLQORLAGREG--- 182
DB 123 QNAIAGLTVNLAKIAKRVLAQPTAFKAETVGNVPRKDSLELDPQRFKARATAPY 182
QY 183 PVRVLVGGSGGARILNQTPQVAAKLGDSVIIHQSGKGSQSQVQAYAEAGQPHKVT 242
DB 183 PLNLVGGSGGARINQTPVAKALGNALFIRHQAGKGLRTISDVYKQADNV--SVT 240
QY 243 EFIDMAAAYAVADVVCRSGLTVSIIAAGLPALFPVPHQKDRQYVNLPLEKAGAA 302
DB 241 EFIDMAEAYNADLVICRSGLTVSIIAAGLPALFPVPHQKDRQYVNLPLEKAGAA 300
QY 303 KIIPQPLSDVAVANTLAGW--SRETLITWAEARAAASIPDAERVAEVSRRVAR 355
DB 301 IIVEQPOFTAEINLNLIOPLKROKLTETWAKATKATKAAQVAVETIIEYSK 355

RESULT 16
MURG_BUCAL STANDARD; PRT; 354 AA.
AC Q8X9T4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.227) (Undecaprenyl-PP-MurNac-pentapeptide-UDP-GlcNac
DE transferase)
GN MURG OR BUS210.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=98794;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22084549; PubMed=12089438;
RA Tanas I., Klasson L., Canback B., Naeslund A.K., Eriksson A.-S.,
RA Wernegreen J.J., Sandstrom J.P., Moran N.A., Andersson S.G.E.;
RT "50 million years of genomic stasis in endosymbiotic bacteria";
RL Science 296:2376-2379 (2002).
CC -!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNac
CC subunit on undecaprenyl-pyrophosphoryl-MurNac-pentapeptide (lipid
CC intermediate I) to form undecaprenyl-pyrophosphoryl-MurNac-
CC (pentapeptide/GlcNac (lipid intermediate II) (By similarity).
CC -!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac(oyl-L-Ala-
CC gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP +
CC GlcNac-(1->4)-Mur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
CC diphosphoundecaprenol.
CC -!- PATHWAY: Peptidoglycan biosynthesis; last step.
CC -!- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG
CC subfamily.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE014097; AAM67773.1; -.
DR FAYAP; MF_00033; -; 1.
DR InterPro; IPR007235; Glyco_tran_28.C.
DR InterPro; IPR004276; Glyco_trans_28.
DR InterPro; IPR006009; MurG.
DR Pfam; PF04101; Glyco_tran_28.C; 1.
DR Pfam; PF03033; Glyco_transf_28; 1.
DR TIGRFAMS; TIGR01133; murG; 1.
DR Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;
KW Peptidoglycan synthesis; Complete proteome.
SQ SEQUENCE 354 AA; 39558 MW; 848B5F42605F2852 CRC64;

Query Match 50.2%; Score 945; DB 1; Length 354;
Best Local Similarity 48.1%; Pred. No. 4.7e-62;
Matches 169; Conservative 79; Mismatches 103; Indels 0; Gaps 0;

QY 7 KRLVMAGGTHGVFPGGLAVAHLMAGQWVRVLGTADRMADLVPHKGIEIDFIRISGL 66
DB 4 KRILLAGSGGTHGVFPGGLTIAKHLIKKGDINWIGTKNIESEIIPKCNIKIHIKQGL 63
QY 67 RKGKIKALIAAPLIFINAWQARAIMKAYKDPVVLGMGVSGPGGLAAWSLGPVLHE 126
DB 64 RNSLNLNLTIPNLVNSYLQVRKLIKWIPIIILGMGVSGPGGLAANSCKIPFILHE 123
QY 127 QNGTAGLTKNLARIATKVMQAPGAPNAEVGNVPTDVLALPLQORLAGREGPVRV 186
DB 124 QNKIAGITNKLKSKISTKNQAFSGTLLNABIVGNPIKNIIDIPPIKRFKNRKGPLRI 183
QY 187 LVVGGSGARILNQTPQVAAKLGDSVIIHQSGKGSQSQVQAYAEAGQPHKVTBEID 246
DB 184 LIIGSGGASIFNKLKIPFLQEKALIIHQSGNNDLQTKRKYKYSTTKHIVSFIK 243
QY 247 DMAAAYAVADVVCRSGLTVSIIAAGLPALFPVPHQKDRQYVNLPLEKAGAAKIE 306
DB 244 NTAAYEWADIIISRSGLTVSEITVGLGAIFFYPHKKDQYVNLADLENNGAAKIE 303
QY 307 QPLSDVAVANTLAGWSRETLITWAEARAAASIPDAERVAEVSRRVARAL 357
DB 304 QSMFAELIKLNSLNREKUFIMAKKAYSIGIRNSTSKISKIHDVSNKI 354

RESULT 17
MURG_BUCAL STANDARD; PRT; 354 AA.
ID MURG_BUCAL
AC P57311;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.227) (Undecaprenyl-PP-MurNac-pentapeptide-UDP-GlcNac
DE transferase).
GN MURG OR BU216.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. APS";
RL Nature 407:81-86 (2000).
CC -!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNac
CC subunit on undecaprenyl-pyrophosphoryl-MurNac-pentapeptide (lipid
CC intermediate I) to form undecaprenyl-pyrophosphoryl-MurNac-

```
CC (pentapeptide)GlcNAc (lipid intermediate II) (By similarity).
CC -!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac(oyl-L-Ala-
CC Gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP +
CC GlcNAc-(1->4)-Mur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
CC diphosphoundecaprenol.
CC -!- PATHWAY: Peptidoglycan biosynthesis; last step.
CC -!- SUBCELLULAR LOCATION: Membrane-associated (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AP001118; BAB12932.1; -.
CC HSSP: P17443; 1FOK.
CC HAMAP: MF_00033; -.
CC InterPro: IPR007235; Glyco_tran_28_C.
CC InterPro: IPR004276; Glyco_tran_28.
CC InterPro: IPR006009; MurG.
CC Pfam: PF04101; Glyco_tran_28_C; 1.
CC Pfam: PF03033; Glyco_transf_28; 1.
CC TIGRFAMs: TIGR01133; murG; 1.
CC Transferase: Glycosyltransferase; Cell division; Cell wall; Membrane;
CC Peptidoglycan synthesis; Complete proteome.
CC KW Peptidoglycan synthesis; Complete proteome.
CC SQ SEQUENCE 354 AA; 39499 MW; 70F23A1294E6B3A6 CRC64;
CC
CC Query Match 49.9%; Score 939; DB 1; Length 354;
CC Best Local Similarity 48.4%; Pred. No. 1.3e-61;
CC Matches 169; Conservative 80; Mismatches 100; Indels 0; Gaps 0;
CC
CC QY 7 KRLVMAGTGGHVPGLAVAHLMQAQGVRLGTADRMADLVPKHGIEIDFIRISGL 66
CC DB 4 KXIIIMAGSGGHVFPFGTIARYITEKGLVNWLTGNTKSIESRIIPKYGIKHVISIKGL 63
CC
CC QY 67 RGKGIKALIAAPLRIFENAWROARAIMKAYKPDVVLGNGVYSGPGGLAAMSLGIPVYLHE 126
CC DB 64 RNTSLKNLIISPIYLTRAYAYVKKIITWSPDIVLGMGVYSGPGVASWNCNIPLLLHE 123
CC
CC QY 127 QNGIAGLTNKLARIATKNQAEPGAFNAEVNPNVYRTDVLALPLPQRLAGREGPVRY 186
CC DB 124 QNKIAGITNKLWLSRISTKNQASPGVLNRAEVNPNVQSIIRKPNINRPNKNTGLRV 183
CC
CC QY 187 LVVGSQGARILNOTMPOVAALGDSVVIWHQSGKSGSQSQVEQAYABAGOPQHKVTEFID 246
CC DB 184 LVIGSQSSSILNRLPEVSPFLKEKLIIFWHTGNYELEKTKKYNKLRNLQNLTISFIK 243
CC
CC QY 247 DMAAYAWADVVCSSGALTSETAAAGLPALEFPVFFQHKDQQQVWALPLEKAGAAKIE 306
CC DB 244 NIASAYEADLTICSGALTSETISIVGLGAIPTYPYPHKQQRHNAEDLELIGAAKIID 303
CC
CC QY 307 QPLSVDAVANTLWAGSRETLITWAERARAASIPDATERVANEVSRAVAR 355
CC DB 304 QSNLTKLVINLSLDKDLFIMAKKAHSLGVDRDAIFNIFNVINKISK 352
CC
CC RESULT 18
CC MURG_CANEF
CC ID MURG_CANEF STANDARD; PRT; 360 AA.
CC AC Q7U346;
CC DT 15-MAR-2004 (Rel. 43, Created)
CC DT 15-MAR-2004 (Rel. 43, Last sequence update)
CC DT 15-MAR-2004 (Rel. 43, Last annotation update)
CC DE UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)
CC DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
CC DE (EC 2.4.1.227) (Undecaprenyl-pp-MurNAc-pentapeptide-UDP-GlcNAc
CC transferase).
CC GN MURG OR BFU142.
CC OS Candidatus Blochmannia floridanus.
```

```
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; ant endosymbionts; Candidatus Blochmannia.
OX NCBI_TaxID=203907;
RN SEQUENCE FROM N.A.
RX MEDLINE=22784745; PubMed=12886019;
RA Gil R., Silva F.J., Zientz E., Delmotte F., Gonzalez-Candelas F.,
RA Latorre A., Rausell C., Kamerbeek J., Gadau J., Hoelldobler B.,
RA van Ham R.C.H.J., Gross R., Moya A.;
RT "The genome sequence of Blochmannia floridanus: comparative analysis
RT of reduced genomes.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:9388-9393(2003).
CC -!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNAc
CC subunit on undecaprenyl-pyrophosphoryl-MurNAc-pentapeptide (lipid
CC intermediate I) to form undecaprenyl-pyrophosphoryl-MurNAc-
CC (pentapeptide)GlcNAc (lipid intermediate II) (By similarity).
CC -!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac(oyl-L-Ala-
CC gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP +
CC GlcNAc-(1->4)-Mur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
CC diphosphoundecaprenol.
CC -!- PATHWAY: Peptidoglycan biosynthesis; last step.
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: BX248584; CAD83663.1; -.
CC HAMAP: MF_00033; -.
CC InterPro: IPR004276; Glyco_trans_28.
CC InterPro: IPR007235; Glyco_tran_28_C.
CC InterPro: IPR006009; MurG.
CC Pfam: PF03033; Glyco_transf_28; 1.
CC Pfam: PF04101; Glyco_tran_28_C; 1.
CC TIGRFAMs: TIGR01133; murG; 1.
CC Transferase: Glycosyltransferase; Cell division; Cell wall; Membrane;
CC Inner membrane; Peptidoglycan synthesis; Complete proteome.
CC KW Inner membrane; Peptidoglycan synthesis; Complete proteome.
CC SQ SEQUENCE 360 AA; 40579 MW; DB0BE4539AD04F46 CRC64;
CC
CC Query Match 47.8%; Score 900; DB 1; Length 360;
CC Best Local Similarity 50.7%; Pred. No. 9.4e-59;
CC Matches 180; Conservative 64; Mismatches 107; Indels 4; Gaps 3;
CC
CC QY 2 MSGQGKRLVMAGTGGHVPGLAVAHLMQAQGVRLGTADRMADLVPKHGIEIDFI 61
CC DB 1 MNGSVQTIIMIAGTGCHIFPGLSVARYLMNHGYKVWVIGSKDRIESELVPPVYNIDIKYI 60
CC
CC QY 62 RISGLEKGI-KALIAAPLRIFENAWROARAIMKAYKPDVVLGNGVYSGPGGLAAMSLGI 120
CC DB 61 CIOGLRGKIYQKLIITLLFLIFAMYSQSPFKIRCKWPKDIVLNGVYSGPSSSLVAVLYGI 120
CC
CC QY 121 PVLVHEQNGIAGLTNKLARIATKNQAEPGAFNAEVNPNVYRTDVLALPLPQRLAGR 180
CC DB 121 PVIIEHQNRIMGLTNRYVSRFAKKIILQGFPNTVNGAITSNPLRVSEILSPDPVHLEGR 180
CC
CC QY 181 EGPVRVLVVGSGGARILNOTMPOVAALGDSVVIWHQSGKSGSQSQVEQAYAE--AGQPQ 238
CC DB 181 TGPVRLVVGSGGTSGSFINKVPEVFGKLFGLKLIWHQSGKSGKGFNDTIQAYKKLHCNSN 240
CC
CC QY 239 HKYTEFTDDMAAYAWADVVCSSGALTSETAAAGLPALEFPVFFQ-HKDRQQVWALPLE 297
CC DB 241 YKVVPFDNWAHAYSADVVIISGALMSEISVGLPAIFVFPFNKDYQQVWNAFQLV 300
CC
CC QY 298 KAGAAKIEQPLSVDAVANTLWAGSRETLITWAERARAASIPDATERVANEVS 352
CC DB 301 KSGSALIIIEQERTSDYVSIIILGNMNRKVLNNAIILSKSLEMSNATQLVAQTVMR 355
```



```
RESULT 19
MURG_SHEON
ID MURG_SHEON STANDARD; PRT; 362 AA.
AC Q8CX35;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.227) (Undecaprenyl-PP-MurNAC-pentapeptide-UDP-GlcNAc GlcNAc
DE transferase).
DE MURG OR S04219.
GN Shewanella oneidensis.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadales; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Nealeson K.H., Fraser C.M.;
RT Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis";
RL Nat. Biotechnol. 20:1118-1123 (2002).
CC -!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNAc
CC subunit on undecaprenyl-pyrophosphoryl-MurNAC-pentapeptide (lipid
CC intermediate I) to form undecaprenyl-pyrophosphoryl-MurNAC-
CC (pentapeptide)GlcNAc (lipid intermediate II) (By similarity).
CC -!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac(Oyl-L-Ala-
CC gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP +
CC GlcNAc-(1->4)-Mur2Ac(Oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
CC diphosphoundecaprenol.
CC -!- PATHWAY: Peptidoglycan biosynthesis; last step.
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE015855; AAN57191.1; -
CC TIGR; SF04219; -
CC HAMAP; MF_00033; -; 1.
CC InterPro; IPR004276; Glyco_trans_28.
CC InterPro; IPR007235; Glyco_tran_28_C.
CC InterPro; IPR006009; MurG.
CC Pfam; PF03033; Glyco_transf_28; 1.
CC Pfam; PF04101; Glyco_tran_28_C; 1.
CC TIGRFAMs; TIGR01133; murG; 1.
CC Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;
CC Inner membrane; Peptidoglycan synthesis; Complete proteome.
CC SEQUENCE 362 AA, 38380 MW, 061728AB95578F8F CRC64;
Query Match 47.2%; Score 887; DB 1; Length 362;
Best Local Similarity 52.5%; Pred. No. 8.4e-58;
Matches 186; Conservative 60; Mismatches 98; Indels 10; Gaps 6;
QY 2 MSGGKGLMNVAGTGGHVFPLGAVAHLMQAQGVWGLGTADRMEDADLPKKGIEIDFI 61
DB 1 MTDAGRLNVNAGTGGHVFPLGAVAHLMQAQGVWGLGTADRMEDADLPKKGIEIDFI 60
```

```
QY 62 RISGLRGKGIKALIAAPLRIFNARQARAIMKAYKPDVILMGVYSGPGGLAAMSLGIP 121
DB 61 DIKVGRLVRLKLAAPFKVRSILQAKVIAEFKPDVILMGVYSGPGGLAAMSLGIP 120
QY 122 VYLHEQNGIAGLTNKLARIATKYNOABPGAPP--NAEVGNPVRTDVLPLPQORLAG 179
DB 121 IYLHEQNAIPGWTNKLISASQVLCAPKNTFTQVKAKVGNPVRRELIALGGPFKQTA- 179
QY 180 REGPVRLVWGSOGARILNQTMTQVAAKLG--DSVIIWHQSGKSGSQSVQEAQVAGQP 237
DB 180 -DEAKLVLVGSGLGAKVFNLMFEVVAALSKQSIIVHGVKDNLAGVKSAYQQQGD 238
QY 238 QH-KVTEFIDDMAAYAWADVVCRSGALTYSEIAAAGLPALFVFPFHQK-DRQYVWNAALP 295
DB 239 GGVVAEFDIDMEAYRWADVVLCRAGALTVELAAVGLPSILVFPYFAVVDHQTNAQV 298
QY 296 LEKAGAAKIIIEPQLSVDAVAN--TLAGSRETLLTWAERARAAASIPATERVA 347
DB 299 LVEAGAAFLPQAILDVNKLVSQIQLLANDRAELARMGQRARDVAVLDATEQVA 352
RESULT 20
MURG_RALSO
ID MURG_RALSO STANDARD; PRT; 365 AA.
AC Q8XVI7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.227) (Undecaprenyl-PP-MurNAC-pentapeptide-UDP-GlcNAc GlcNAc
DE transferase).
DE MURG OR RSC2844 OR RS00261.
GN Ralstonia solanacearum (pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GMI1000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,
RA Caspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Signier P., Inebault P., Whalen M., Wincker P., Levy M.,
RA Weisenbach J., Boucher C.A.;
RL "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502 (2002).
CC -!- FUNCTION: Cell wall formation. Catalyzes the transfer of a GlcNAc
CC subunit on undecaprenyl-pyrophosphoryl-MurNAC-pentapeptide (lipid
CC intermediate I) to form undecaprenyl-pyrophosphoryl-MurNAC-
CC (pentapeptide)GlcNAc (lipid intermediate II) (By similarity).
CC -!- CATALYTIC ACTIVITY: UDP-N-acetylglucosamine + Mur2Ac(Oyl-L-Ala-
CC gamma-D-Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol = UDP +
CC GlcNAc-(1->4)-Mur2Ac(Oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-Ala)-
CC diphosphoundecaprenol.
CC -!- PATHWAY: Peptidoglycan biosynthesis; last step.
CC -!- SUBCELLULAR LOCATION: Inner membrane-associated (By similarity).
CC -!- SIMILARITY: Belongs to the glycosyltransferase family 28. MurG
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL646072; CAD16551.1; -
CC HAMAP; MF_00033; -; 1.
```


DR InterPro; IPR007235; Glyco_tran_28.C.
 DR InterPro; IPR004276; Glyco_tran_28.
 DR InterPro; IPR006009; MurG.
 DR Pfam; PF04101; Glyco_tran_28.C; 1.
 DR Pfam; PF03033; Glyco_tran_28; 1.
 DR TIGRFAMs; TIGR01133; murG; 1.
 KW Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;
 KW Inner membrane; Peptidoglycan synthesis; Complete proteome.
 SQ SEQUENCE 365 AA; 3796 MW; BB9DD2C0CFIA88A CRC64;

Query Match 46.2%; Score 869.5; DB 1; Length 365;
 Best Local Similarity 51.0%; Pred.No.1.6e-56;
 Matches 181; Conservative 55; Mismatches 114; Indels 5; Gaps 3;

QY 7 KRLMTMAGTGGHVPFGLVAHHLMAQGVRLGTADRMEDLVPKHGIEIDFIRISGL 66
 DB 11 RTLVMTAGTGGHIFPALSVARLLAARGQVVLGNASGMEGLVPKHPLESVRFGV 70
 QY 67 RGKGIKALIAAPLRFNARQARAIMKAYKPDVVLMGMYVSGGLAANSLGIPVVLHE 126
 DB 71 RGKGLVTRKFLPLNLLRAFMQSLGVRRVRPNVVLGMGGYITFPFGNMVLLGAPLVLHE 130
 QY 127 QNGIAGLTNKLARIATKMOAEFGAPNAEVCNPNVDTVLALPLPQRLAGREGPVRV 186
 DB 131 QNSIAGLANRVLARVADRVLCAPFGALPFGAEWGNPIRADLAALPSQARYAERSGPIRV 190
 QY 187 LVVGGSGQARIINOTMPQVAAKLGDSV--IITHQSGKSGSQSQVEQAYAEAG--QPOHKVT 242
 DB 191 LVVGGSLGAALNDVVPALALLPADTPIVIHQAGAKQIDTLRANYAAVIGIDETHAQAV 250
 QY 243 EFIDMAAAYAWADVVCRSGLTVSEIAAAGLPALFVPPQHK-DRQYNNALPLEKAGA 301
 DB 251 PFIDMAAAYAAQADLVI CRAGAMTVSEVAAAAGVAALFVPPFHAVDDHQTNNARFLSERGA 310
 QY 302 AKIIEPOLSVDAVANTLAGWSRETLTMAERARAASIPDATERVANEVSRVAA 356
 DB 311 ALLVPQSLGPASLADTLASLTQAQADMAKAKAEQARPEAEAEVADICVAAAEA 365

Search completed: June 7, 2004, 07:13:42
 Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 7, 2004, 07:02:25 ; Search time 45 Seconds

(without alignments)

2552.192 Million cell updates/sec

Title: US-09-829-275-1

Perfect score: 1881

Sequence: 1 MMSGQKRLVMAGGTGGHV.....RVANEVSRVARELHHHHH 364

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

SPTREMBL_25:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_todent:*

12: sp_virus:*

13: sp_vertibrate:*

14: sp_unclassified:*

15: sp_rvirus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|---------------------|
| 1 | 1793 | 95.3 | 355 | 16 | Q83MM4 shigella fl |
| 2 | 1021 | 54.3 | 355 | 16 | Q7U336 haemophilus |
| 3 | 900 | 47.8 | 360 | 16 | Q7U346 candidatus |
| 4 | 887 | 47.2 | 362 | 16 | Q8CX35 shewanella |
| 5 | 794 | 42.2 | 357 | 16 | Q7W4B4 bordetella |
| 6 | 793 | 42.2 | 357 | 16 | Q7WFS2 bordetella |
| 7 | 791 | 42.1 | 357 | 16 | Q7VUC3 bordetella |
| 8 | 787 | 41.8 | 358 | 16 | Q820X3 coxiella bu |
| 9 | 764 | 40.6 | 357 | 16 | Q82VS3 nitrosomona |
| 10 | 717 | 38.1 | 356 | 16 | Q87WY5 pseudomonas |
| 11 | 683 | 36.3 | 359 | 16 | Q88N76 pseudomonas |
| 12 | 506 | 26.9 | 410 | 16 | Q7VEP8 mycobacteri |
| 13 | 496.5 | 26.4 | 379 | 16 | Q8CY39 bruceella su |
| 14 | 480.5 | 25.5 | 360 | 16 | Q8FNU0 corynebacte |
| 15 | 476 | 25.3 | 372 | 16 | Q8A2S8 bacteroides |
| 16 | 461.5 | 24.5 | 363 | 16 | Q820F6 streptomyce |

| | | | |
|-----|----|--------|---------------------|
| 379 | 2 | Q8GBA0 | Q8gba0 heliobacill |
| 393 | 16 | Q8CY50 | Q8cy50 bifidobacte |
| 356 | 16 | Q83HK1 | Q83hk1 tropheryma |
| 356 | 16 | Q820Y4 | Q820y4 tropheryma |
| 363 | 16 | Q88V81 | Q88v81 lactobacill |
| 311 | 2 | Q7RQK0 | Q7rdk0 caulobacter |
| 361 | 16 | Q7V466 | Q7v466 prochloroco |
| 318 | 16 | Q9RWPO | Q9rwp0 deinococcus |
| 358 | 16 | Q7U3U6 | Q7u3u6 synechococc |
| 366 | 16 | Q89FU7 | Q89fu7 bradyrhizob |
| 357 | 16 | Q8R5N5 | Q8r5n5 fusobacteri |
| 357 | 16 | Q7VDZ2 | Q7vdz2 prochloroco |
| 357 | 16 | Q8CUL4 | Q8cul4 oceanobacil |
| 369 | 16 | Q893R7 | Q893r7 clostridium |
| 431 | 16 | Q9C9T8 | Q9c9t8 arabidopsi |
| 364 | 16 | Q7V388 | Q7v388 prochloroco |
| 352 | 16 | Q812T8 | Q812t8 bacillus ce |
| 352 | 16 | Q81J66 | Q81j66 bacillus an |
| 358 | 16 | Q820E0 | Q820e0 chlamydophi |
| 350 | 16 | Q7U322 | Q7u322 halicobacte |
| 354 | 16 | Q812Y1 | Q812y1 bacillus ce |
| 358 | 16 | Q8F4J1 | Q8f4j1 leptospira |
| 352 | 16 | Q8DQJ1 | Q8dqm1 streptococc |
| 653 | 10 | Q8LQD0 | Q8lqd0 oryza sativ |
| 357 | 16 | Q8CMM3 | Q8cmm3 staphylococ |
| 360 | 2 | Q9FB02 | Q9fb02 streptococc |
| 358 | 16 | Q8E6P0 | Q8eep0 streptococc |
| 358 | 16 | Q8CX15 | Q8cx15 streptococc |
| 361 | 16 | Q8DVE2 | Q8dve2 streptococc |
| 356 | 16 | Q8TYD0 | Q8tyd0 methanopyru |
| 388 | 16 | Q7TY01 | Q7ty01 mycobacteri |
| 420 | 16 | Q33282 | Q33282 mycobacteri |
| 397 | 16 | Q825U1 | Q825u1 streptomyce |
| 425 | 2 | Q83W14 | Q83w14 streptomyce |
| 396 | 2 | Q9RQR5 | Q9rrq5 streptomyce |
| 402 | 16 | Q88GN0 | Q88gn0 rhizobium l |
| 392 | 2 | Q8KNF2 | Q8knf2 micromonosp |
| 392 | 16 | Q49841 | Q49841 mycobacteri |
| 402 | 2 | Q9F8U7 | Q9f8u7 streptomyce |
| 427 | 16 | Q82VF2 | Q82vf2 rhizobium m |
| 388 | 2 | Q9RPA1 | Q9rpa1 streptomyce |
| 382 | 2 | Q9R2F9 | Q9rf29 streptomyce |
| 346 | 16 | Q92Y00 | Q92y00 rhizobium m |
| 525 | 10 | Q93115 | Q93115 cucumis sat |
| 112 | 2 | Q54027 | Q54027 porphyromon |
| 433 | 16 | Q73403 | Q73403 synecocyst |
| 522 | 10 | Q9SM44 | Q9sm44 spinacia ol |
| 379 | 2 | Q93H13 | Q93h13 streptomyce |
| 378 | 16 | Q7WP06 | Q7wp06 bordetella |
| 378 | 16 | Q7W1A2 | Q7w1a2 bordetella |
| 418 | 16 | Q8PJG6 | Q8pjg6 xanthomonas |
| 378 | 16 | Q7VUE9 | Q7vue9 bordetella |
| 519 | 17 | Q27324 | Q27324 methanobact |
| 376 | 2 | Q8KND7 | Q8knd7 micromonosop |
| 535 | 10 | Q9FZL3 | Q9fzl3 nicotiana t |
| 237 | 17 | Q27683 | Q27683 methanobact |
| 390 | 2 | Q9ZGB8 | Q9zgb8 streptomyce |
| 530 | 10 | Q9FZL4 | Q9fzl4 glycine max |
| 559 | 16 | Q8HZU4 | Q8hzu4 pseudomonas |
| 388 | 16 | Q8UCA6 | Q8uca6 agrobacteri |
| 390 | 2 | Q8GHC2 | Q8ghc2 streptomyce |
| 411 | 16 | Q9RVF3 | Q9rvf3 deinococcus |
| 444 | 16 | Q8P450 | Q8p450 xanthomonas |
| 370 | 16 | Q9PEZ9 | Q9pez9 xylella fas |
| 391 | 2 | Q9RP99 | Q9rp99 streptomyce |
| 357 | 16 | Q87PP1 | Q87pp1 vibrio para |
| 352 | 2 | P74819 | P74819 shingomona |
| 428 | 2 | O88003 | O88003 bordetella |
| 430 | 16 | Q7WR12 | Q7wr12 bordetella |
| 430 | 16 | Q7W234 | Q7w234 bordetella |
| 463 | 2 | Q50458 | Q50458 mycobacteri |
| 360 | 16 | Q92VH0 | Q92vh0 rhizobium m |
| 435 | 16 | Q9CD88 | Q9cd88 mycobacteri |

90 116.5 6.2 438 16 Q9CD91 mycobacteri
91 116 6.2 465 10 Q9S193
92 115.5 6.1 91 2 Q93SU1
93 115.5 6.1 390 2 Q9ALM8
94 115.5 6.1 428 16 P95134
95 115.5 6.1 1049 16 Q891T6
96 115 6.1 388 16 Q9WZ90
97 115 6.1 428 16 Q45374
98 115 6.1 449 16 P95130
99 115 6.1 449 16 Q7TXJ4
100 115 6.1 1273 17 Q9YCA1

ALIGNMENTS

RESULT 1
Q83MN4 PRELIMINARY; PRT; 355 AA.
AC Q83MN4;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE UDP-N-acetylglucosamine:N-acetylmuramyl- (pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase.
GN MURG OR SP0087 OR S0089.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RC MEDLINE=2272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
Yu J.;
RA "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
through comparison with genomes of Escherichia coli K12 and O157";
RL Nucleic Acids Res. 30:4432-4441 (2002).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RC MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
Schwartz D.C., Blattner F.R.;
RA "Complete genome sequence and comparative genomics of Shigella
flexneri serotype 2a strain 2457T";
RL Infect. Immun. 71:2775-2786 (2003).
DR EMBL; AE015046; AAN41752.1; -.
DR EMBL; AE016978; AAP15633.1; -.
DR GO; GO:0019866; C:inner membrane; IEA.
DR GO; GO:0016758; F:transferase activity, transferring hexosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0030259; P:lipid glycosylation; IEA.
DR GO; GO:0019277; P:UDP-N-acetylglucosamine biosynthesis; IEA.
DR InterPro; IPR004276; Glyco_trans_28.
DR InterPro; IPR007235; Glyco_tran_28_C.
DR InterPro; IPR006009; Murg.
DR Pfam; PF03033; Glyco_transf_28; 1.
DR Pfam; PF04101; Glyco_tran_28_C; 1.
DR TIGRFAMs; TIGR01133; murg; 1.
KW Transferase; Complete proteome.
SQ SEQUENCE 355 AA; 37812 MW; 73407776C2B1504C CRC64;

Query Match 95.3%; Score 1793; DB 16; Length 355;
Best Local Similarity 98.6%; Pred. No. 1.5e-119;
Matches 350; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 MSGCKRLVMAGCTGGHVPGLAVAHHLMAQGWVWMLGTADRVADLVPKHGIEIDFI 61
DB 1 MSGCKRLVMAGCTGGHVPGLAVAHHLMAQGWVWMLGTADRVADLVPKHGIEIDFI 60
QY 62 RISGLRGKIGIKALIAAPLIRIFNARQARAIMKAYKDPVILMGVYSGPGGLAAWSLIGIP 121
DB 61 RISGLRGKIGIKALIAAPLIRIFNARQARAIMKAYKDPVILMGVYSGPGGLAAWSLIGIP 120
QY 122 VVLEHONGIAGLTNKLWARIATKMQAEPGAFPAEVEVGNPVRTDVLALPLPQORLAGRE 181
DB 121 VVLEHONGIAGLTNKLWARIATKMQAEPGAFPAEVEVGNPVRTDVLALPLPQORLAGRE 180
QY 182 GPRVVLVVGSGQGARILNQTMPQVAAKLGDSVITWHQSGKGSQSQVEQAYAEAGQPQHKV 241
DB 181 GPRVVLVVGSGQGARILNQTMPQVAAKLGDSVITWHQSGKGSQSQVEQAYAEAGQPQHKV 240
QY 242 TEFDIDMAAAYAWADVVCRCGALTVSEIAAAGLPALFVFPQHKDROQYNNALPLEKAGA 301
DB 241 TEFDIDMAAAYAWADVVCRCGALTVSEIAAAGLPALFVFPQHKDROQYNNALPLEKAGA 300
QY 302 AKIIEQPOLSDVAVANTLAGRSRETLTMAERARAASIPDATERVANEVSRAA 356
DB 301 AKIIEQPOLSDVAVANTLAGRSRETLTMAERARAASIPDATERVANEVSRAA 355
RESULT 2
Q7U336 PRELIMINARY; PRT; 355 AA.
AC Q7U336;
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase.
GN MURG OR H00824.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=35000HP / ATCC 700724;
RC Munson R.S. Jr., Ray W.C., Mahairas G., Sabo P., Mungur R.,
Johnson L., Nguyen D., Wang J., Forst C., Hood L.;
RA "The complete genome sequence of Haemophilus ducreyi";
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017153; AAP95721.1; -.
KW Transferase; Complete proteome.
SQ SEQUENCE 355 AA; 38478 MW; EB74CBE2BFB6574 CRC64;

Query Match 54.3%; Score 1021; DB 16; Length 355;
Best Local Similarity 57.5%; Pred. No. 1.3e-64;
Matches 204; Conservative 51; Mismatches 92; Indels 8; Gaps 3;
QY 7 KRLVMAGCTGGHVPGLAVAHHLMAQGWVWMLGTADRVADLVPKHGIEIDFIRISGL 66
DB 3 KLLIMAGCTGGHVPGLAVAHHLMAQGWVWMLGTADRVADLVPKHGIEIDFIRISGL 62
QY 67 RGKGIKALIAAPLIRIFNARQARAIMKAYKDPVILMGVYSGPGGLAAWSLIGIPVVLHE 126
DB 63 KKGGLALIKAPFTTLKAVLQALNIIKKYRPDVLGMGVYSGPGGLAAWSLIGIPVVLHE 122
QY 127 QNGIAGLTNKLWARIATKMQAEPGAFPAEVEVGNPVRTDVLALPLPQORLAGRE---- 182
DB 123 QNAIAGLTNKLWARIATKMQAEPGAFPAEVEVGNPVRTDVLALPLPQORLAGRE---- 182
QY 183 PVRVLVVGSGQGARILNQTMPQVAAKLGDSVITWHQSGKGSQSQVEQAYAEAGQPQHKV 242
DB 183 PLNVLVVGSGQGARILNQTMPQVAAKLGDSVITWHQSGKGSQSQVEQAYAEAGQPQHKV 240
QY 243 EFIDDMAAAYAWADVVCRCGALTVSEIAAAGLPALFVFPQHKDROQYNNALPLEKAGA 302
DB 241 EFIDDMAAAYAWADVVCRCGALTVSEIAAAGLPALFVFPQHKDROQYNNALPLEKAGA 300

| | |
|----------|---|
| RA | Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A., |
| RA | Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I., |
| RA | Chillingworth T., Collins M., Cronin A., Davis P., Doggett J., |
| RA | Fellwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K., |
| RA | Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C., |
| RA | Rabinowitch E., Rutter S., Sanders M., Saunders D., Seeger K., |
| RA | Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K., |
| RA | Unwin L., Whitehead S., Barrell B.G., Maskell D.J.; |
| RT | "Comparative analysis of the genome sequences of Bordetella pertussis, |
| RT | Bordetella parapertussis and Bordetella bronchiseptica."; |
| RL | Nat. Genet. 35:32-40(2003). |
| RL | EMLB; BX60449; CAE34562.1; -- |
| KW | Glycosyltransferase; Transferase; Complete proteome. |
| SQ | SEQUENCE 357 AA; 37849 MW; BC5D8687EE0F48AF CRC64; |
| | |
| | Query Match 42.2%; Score 793; DB 16; Length 357; |
| | Best Local Similarity 48.3%; Pred. No. 2.2e-48; |
| | Matches 169; Conservative 55; Mismatches 122; Indels 4; Gaps |
| QY | 9 LMVAGGTGGHVFPGGLAVAHILMAQGVOVRWLGTADRMEDALVPKHGIEIDFIRISGLRG 65 |
| DB | : |
| QY | 6 ILIIMAGTGGHIMPGLAFAVEVLREGRWVLNLDNPDKMEGRLLPVGRIELVPLRFQGVRG 65 |
| DB | : |
| QY | 69 KGKIALIAAPLIFINAWQARAIMKAYKPDVVVLGMGVGYSGPGGLAANSLGIPVVLHQON 12 |
| DB | : |
| QY | 66 RGAALLKLFPFLARACACAQRRLADIIPDDVVLGMGVGYVAFPGGYAAIRRTLPLVHHQN 12 |
| DB | : |
| QY | 129 GTAGLTNKWLARIATKWQAEPGFAPFNVECVGNPVRTDLALPLPOORLAGREGPVRLV 18 |
| DB | : |
| QY | 126 AVAGTANRWLARLRVLSGGPPGVLPGRGALGNPVRLDALCALPEPAERYAGRSGLRVLV 18 |
| DB | : ~~~~ |
| QY | 189 VGSQCARILNTMPQVAAKLGDSV--IIHQSGKGSSQSVEAYAEAGQPQHKTFFID 24 |
| DB | : ~~~~ |
| QY | 247 DMAAYADVNVVCRSGLATTSEIIAAGLIPALFVEFHQK-DQQYWNALPLEKAGAAKII 30 |
| DB | : ~~~~ |
| QY | 245 DWAGAWAQADLLICRAGAMTVSEVAAAQVAAVLVFPFPAIDHDQTANARFLSDAQAAQLQ 30 |
| DB | : ~~~~ |
| QY | 306 EQPOLSVDAVANTLAGWSRETILLTAERARAASIPDATERVANEYSRVAR 355 |
| DB | : ~~~~ |
| QY | 305 PQASLTTPWLQWLQRTQEQLQAVAGRTHALPRAAAHITADVCEQAAR 354 |
| DB | : ~~~~ |
| RESULT 7 | |
| Q7VUQ3 | PRELIMINARY; PRT; 357 AA. |
| ID | Q7VUQ3 Q7VUQ3 |
| DT | 01-OCT-2003 (TEMBLrel_25, Created) |
| DC | 01-OCT-2003 (TEMBLrel_25, Last sequence update) |
| DT | 01-OCT-2003 (TEMBLrel_25, Last annotation update) |
| DE | UDP-N-acetylglucosamine-N-acetylmuramyl- (EC 2.4.1.-)- |
| OS | MURG OR BP3023 |
| GN | Bordetella pertussis |
| OC | Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; |
| OC | Alcaligenaceae; Bordetella. |
| OX | NCBI_Taxid=520; |
| RN | [1] |
| RP | SEQUENCE FROM N.A. |
| RC | STRAIN=Tohnana I / ATCC BAA-589 / NCTC 13251; |
| RC | MEDLINE=22827954; PubMed=12910271; |
| RA | Parkhill J., Sebaihia M., Preston A., Churcher C., Murphy L.D., Thomson N., |
| RA | Harris D.E., Holden M.T.G., Churcher C., James K., Harris B., Quail M.A., |
| RA | Cerdeno-Tarraga A.M., Temple L., James K., Basham D., Bason N., Cherevach I., |
| RA | Achtman M., Atkin R., Baker S., Basham D., Cronin A., Davis P., Doggett J., |
| RA | Chillingworth T., Collins M., Cronin A., Hauser H., Holroyd S., Jagels K., |
| RA | Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C., |
| RA | Rabinowitch E., Rutter S., Sanders M., Saunders D., Seeger K., |
| RA | Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K., |
| RA | Unwin L., Whitehead S., Barrell B.G., Maskell D.J.; |
| RT | "Comparative analysis of the genome sequences of Bordetella pertussis, |
| RT | Bordetella parapertussis and Bordetella bronchiseptica."; |

Matches 162; Conservative 68; Mismatches 114; Indels 8; Gaps 4

QY 10 MMAGTGGHVFPGGLAVAHLMACGQVRLGTADREADLVPKHGIEIDFIRISGLRGK 69
DB 1 MIMAGTGGHVFPGGLAVARSQANGWEIVLGRTRNGEALVPQHGFSTELINPFLSGK 60

QY 70 GKIALIAPLRIFNWQAARIMKAYDPDVLGMGGTVSPGGLAWSLGIPVLHEONG 129
DB 61 KLSVVLLFWRLAQACQSFILRRQPQVVLGMGGYPALPGIIMAVLKGKELLTHEQR 120

QY 130 IAGLTNKKWLARIATKWAEPCAPFNAE----VVGNPVRTDLALPLPOORLAGREGPVR 185
DB 121 IAGLTNKLAIADRILLAFGALTSPEKTRVTGNPRTETARLPSPPEARVAHTGNLH 180

QY 186 VLWVGSSOGAILNQTPQVAAKLDGS--VIIHQSKGQQSQVEQAYAAEQPHKVTE 243
DB 181 ILWVGSLGAQVLNTVLPIQALSMIPEDQRPVYTHQSCKAHLDALQOAYADHGVGTNLVA- 239

QY 244 FTDMAAAYANADVVCBSGALTSETAAAGLPALEFPFOHK--DRQYNWALPLEKAGAA 302
DB 240 FIENNAHYQDCDLVICRAGALTISELAAGVASILIPPIYAADDHOTANARFLSDYQAA 299

QY 303 KIIEPOLSVDAVANTLAGSRETLLTWAPERARAASIPDATERVANEVSRA 354
DB 300 VLWPQSELTAASLAQWLMTCSRAQLQSMATHARALAMPEAAQTVAECQQLS 351

RESULT 10

Q87WYS PRELIMINARY; PRT; 356 AA.

ID Q87WYS AC Q87WYS;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE UDP-N-acetylglucosamine--N-acetylmuramyl- (pentapeptide)
DE pyrophosphoryl-undecaprenol-N-acetylglucosamine transferase.
GN MURG OR PSPTO4408.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_Taxid=323;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=DC300ard
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA Barry K., Utterback T., Van Aken S., Feldblyum T., Winn W.,
RA Dodson R., DeBooy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,
RA White O., Fraser C., Collier A.;
RT "Complete sequence of Pseudomonas syringae."
RL Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR ENBL; AEO16871; AAC57857.1; --
DR TIGR; PSPTO4408; --
DR GO; GO:0016758; F:transferase activity, transferring hexosyl . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0030259; P:lipoic acid biosynthesis; IEA.
DR InterPro; IPR004276; Glyco.trans.28.
DR InterPro; IPR007235; Glyco.trans.28.C.
DR Pfam; PF03033; Glyco.transf.28; 1.
DR Pfam; PF04101; Glyco.trans.28.C; 1.
DR Complete proteome.
SQ SEQUENCE 356 AA; 3798 MW; 59956CE06DFC52C CRC64;

Query Match 38.1%; Score 717; DB 16; Length 356;
Best Local Similarity 44.1%; Pred.No. 5.6e-43;
Matches 159; Conservative 63; Mismatches 117; Indels 20; Gaps 9

QY 9 LMWAGTGCHVFPGGLAVAHLMACGQVRLGTADREADLVPKHGIEIDFIRISGLRG 68
DB 5 VLIMAGTGCHVFPGGLAVAHLMACGQVRLGTADREADLVPKHGIEIDFIRISGLRG 64

QY 69 KGKIALIAPLRIFNWQAARIMKAYDPDVLGMGGTVSPGGLAWSLGIPVLHEON 128


```
Db 121 LVIHQNARACTANELLVPLSARVCEAFFNTFEASDKRTTGNVPRPELF---MDAQRTP 177
Qy 179 GREGFVRVLVGGSGQAKILNQTWPQVAAKLGDSV--IIWHSQSGSQSQSVQAYAEAGQ 236
Db 178 LGERARLLVLGGSLGAPLKNLKPALRPEVFAALRPEVFOAGKQAFITABRYHEAGV 237
Qy 237 POHKVTEFIDMAAYAWADVVCRSGALTIVSEIAAAGLPLALFPVFOHK-DRQOYNWALP 295
Db 238 AA-QVEPIKDAQYAGWADLVVCRAGALTIVSELAAAGLPSMLVPLPHADDDHQTHNAQY 296
Qy 296 LEKGAAGKIIQFQPSLVDAVANTL--AGWSRETLTWAERARAASIPDATERVANEVSRV 353
Db 297 LAREGAFLPQATTGAQALAERLNEVLMOPEKLVNMGATARRLAKPAATSTVVDICLEV 356
Qy 354 A 354
Db 357 A 357

RESULT 12
Q7VEP8 PRELIMINARY; PRT; 410 AA.
AC Q7VEP8;
DT 01-OCT-2003 (TREMELrel. 25, Created)
DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)
DE pyrophosphoryl-undecaprenol-N-acetylglucosamine transferase Murg (EC 2.4.1.-)
GN MURG OR M2177C.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97; PubMed=12789972;
RX MEDLINE=22709107;
RA Garnier T., Eglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Dutthoy S., Grogdint S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248341; CAD97030.1; --
DR Transferase; Glycosyltransferase; Complete proteome.
SQ SEQUENCE 410 AA; 41858 MW; 347DB45D34A4890B CRC64;

Query Match 26.9%; Score 506; DB 16; Length 410;
Best Local Similarity 37.8%; Pred. No. 6.9e-28;
Matches 143; Conservative 55; Mismatches 148; Indels 32; Gaps 10;

Qy 3 SQGQKRLVMAGGTGGHVPGLVAHLMAGQVQRV--WLGTADRMEDLVFKHGIEIDF 60
Db 32 SADSLSVLGGGTAGHVEPAMAVADALVALDPRVRITAGLTPRGLETRLVPQGYHLEL 91
Qy 61 IRISGLRGKIGKIALIAAPLRFENAWROARIMKAYKPDVLMGMYGYSGPCGLAANSI-- 118
Db 92 ITAVPMRKPQGDRLARLPSRVWRVAREARDVLDDVDADVGVFGGYPALPAYLAARGLPL 151
Qy 119 -----GTPVLVHEQNGIAGLTNKLARIATKVMQAEF-GAPPNAEVVGNFVRTDVLALP 171
Db 152 PPRRRRFPVVIHEANARAGLANRYGAHTADRVLSAVPDSGLRRAEVGVVPRASIALD 211
Qy 172 LPQQLAGR-----EGPVRVLV--GSGGARILNQTWPQVAAKLGDS--VIWHSQSGSQ 224
Db 212 RAVLRAEAFARHFDPDDARLVLLVFGSGQSVLNRSAVSGAAADLAAAGCVLH--AHGPQ 269
Qy 225 QSQEQAYAEAGOPQKVTETFEIDMAAAVADVVVCRSGALTIVSEIAAAGLPLALFPVFOH 284
Db 270 NVLELRRAGQGPVAVVPIYLDREMLAYAADLVICRAGMTVAEVSAGVLPALVYVPLPI 329
```

```
Qy 285 KDRQQYMNALPLEKAGAAKIIIEQPOLSVDAVANTLWAGSRETLTWAERARAASIPDATE 344
Db 330 GNGEQLNALPVNAGGGMVADALTPELVARQVAG-----LLT--DPARLAANTAAAA 382
Qy 345 RV-----ANEVSRVARAL 357
Db 383 RVGHRDAAGQVAAALAV 400

RESULT 13
Q8CV39 PRELIMINARY; PRT; 379 AA.
AC Q8CV39;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase.
GN MURG OR BR1431.
OS Brucella suis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1330 / Biovar 1;
RX MEDLINE=2247741; PubMed=12271122;
RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
RA Daugherty S.C., Debay R.T., Durkin A.S., Kolonay J.F., Madupu R.,
RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
RA Riedmuller S., Tetelin H., Gill S.R., White O., Salzberg S.L.,
RA Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Fraser C.M.;
RT "The Brucella suis genome reveals fundamental similarities between
RT animal and plant pathogens and symbionts."
RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
DR EMBL; AE014438; AAN30344.1; --
DR TIGR; BR1431; --
DR GO; GO:0019866; C:inner membrane; IEA.
DR GO; GO:0016758; F:transferase activity, transferring hexosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0030259; P:lipid glycosylation; IEA.
DR GO; GO:0019277; P:UDP-N-acetylglucosamine biosynthesis; IEA.
DR InterPro; IPR004276; Glyco trans 28.
DR InterPro; IPR007235; Glyco tran_28_C.
DR InterPro; IPR006009; Murg.
DR Pfam; PF03033; Glyco.transf.28; 1.
DR Pfam; PF04101; Glyco.tran_28_C; 1.
DR TIGRFAMs; TIGR01133; murg; 1.
KW Complete proteome.
SQ SEQUENCE 379 AA; 40236 MW; 705EE98D96F30177 CRC64;

Query Match 26.4%; Score 496.5; DB 16; Length 379;
Best Local Similarity 37.0%; Pred. No. 3e-27;
Matches 134; Conservative 53; Mismatches 154; Indels 21; Gaps 10;

Qy 2 MSQGKRLVMAGGTGGHVPGLVAHLMAGQVQRVWLGT---ADRMEDLVFKHGIEI 58
Db 4 LANQGV-IVLAGGTGGLHFPAAALAEHLRARGWDVH-LATDARAQRFVGAFAQDH--V 58
Qy 59 DFTIRISGLRGKIGKIALIAAPLRFENAWROARIMKAYKPDVLMGMYGYSGPCGLAANSI 118
Db 59 HVIRSATIAGRNPVALLKTFWSLQQNLSRKLFRLEKPLVVGFGYPTLPPLPYAASNM 118
Qy 119 GIPVVLHEQNGIAGLTNKLARIATKVMQAEFGAPPNAEVV-GNPFVRTDVLALPLP 173
Db 119 GIPVVLHEQNGIAGLTNKLARIATKVMQAEFGAPPNAEVV-GNPFVRTDVLALPLP 178
Qy 174 QQLAGREGPVRVLVGGSGGARILNQTWPQVAAKLGDS-----SVIIWHSQSGSQSVQ 229
Db 179 PYTPAGKDRFRLLVFGSGQAGFFSQAIPEAAVALLPEHERARLLITQARKDEASAFQ 238
Qy 230 AYAEAGOPQKVTETFEIDMAAAVADVVVCRSGALTIVSEIAAAGLPLALFPVFOHK-DRQ 288
```

```
Db 239 AYKLGVPA-DVAPFFNDMPARMADAHFVIARGASTVSEITVIGFPMVLVPPFHALDHD 297
Qy 289 QYNALPLEKAGAAKITEIQPQSLSDAVANTL--AGWSRETLTMAERARAASIPDAT 346
Db 298 QANAAALAAAGAEVVRQADLSPQRLAEMQLQSAEMPERLEQQAKAAKSVGKPDARLL 357
Qy 347 AN 348
Db 358 AD 359

RESULT 14
Q8ENU0
ID Q8ENU0 PRELIMINARY; PRT; 360 AA.
AC Q8ENU0
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide)
DE Peptidoglycan biosynthesis protein Murg.
GN MURG OR CB2053.
OS Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=152794;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA Kawarabayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA Usuda Y., Sugimoto S.;
RT "The entire genomic sequence of Corynebacterium efficiens YS-314."
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005221; BAC1863.1; -.
DR GO; GO:0019866; C:inner membrane; IEA.
DR GO; GO:0016758; F:transferase activity, transferring hexosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0030259; P:lipid glycosylation; IEA.
DR GO; GO:004276; Glyco trans. 28.
DR InterPro; IPR004276; Glyco trans. 28.
DR InterPro; IPR007235; Glyco trans. 28.
DR InterPro; IPR007235; Glyco trans. 28.
DR InterPro; IPR007235; Glyco trans. 28.
DR InterPro; IPR007235; Glyco trans. 28.
DR Pfam; PF03033; Glyco trans. 28; 1.
DR Pfam; PF04101; Glyco trans. 28 C; 1.
DR TIGRfam; TIGR01133; murG; 1.
DR PROSITE; PS0038; HLH_1; 1.
KW Complete proteome.
SQ SEQUENCE 360 AA; 36543 MW; 1B37ACF96048E0D CRC64;

Query Match 25.5%; Score 480.5; DB 16; Length 360;
Best Local Similarity 35.6%; Pred. No. 3.8e-26;
Matches 132; Conservative 55; Mismatches 137; Indels 47; Gaps 9;

Qy 9 LMWAGGTGGHVPGLAVAHLM--AQGVQVRLGTADRMADLVPKHGIEIDFIRISGLR 67
Db 10 VVAGGCTAGHIEPALVAEALRDGVCATVTALGTARGLETSLVDPDGLRLIEBPVP 69
Qy 68 KGKTKALIAAPLRFNNAWRQARIMKAYKPDVVLGMGVGPGGLAANSLGIPVVLHEQ 127
Db 70 RTPNLDLVKLPFRVAKSLRQARVLTGTGAHVVGFGYVSAPAYLAARSLGIPFFVHEA 129
Qy 128 NGIAGLNKWLARATKVMQAEPAFFNAEVNVPVTRDVLALPLQORLAGREGPV--- 184
Db 130 NARAGMANKLVLGGVGLNATENSGMPGVVGIPR-----RELAGEDATAAE 179
Qy 185 -----RVLVVGGSGGARILNQTMPQVAUKLGD-----SVIIHQSGKGSQQSV 227
Db 180 RGRQTWGLEADRPVTLVTTGSGQSVNSA---VAGALDILLGAGIQVLAHVGK--RNSL 234
Qy 228 EQAYEAGQPOKHYTEITDDMAAAVAVADVVVCRSGALTVEIAAAGLPALFVFPQHKR 287
Db 235 PTA-----RPGYVVPFIEDMQAAVAVADLVVCRSGANTVAETASGIPAIYVPLPHNG 289
```

RESULT 15

```
Q8A258
ID Q8A258 PRELIMINARY; PRT; 372 AA.
AC Q8A258
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase.
GN BT3448.
OS Bacteroides thetaiotaomicron.
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=818;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI-5482 / ATCC 29148;
RX MEDLINE=22550858; PubMed=12663928;
RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
RA Chiang H.C., Hooper L.V., Gordon J.I.;
RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis."
RL Science 299:2074-2076(2003).
DR EMBL; AB016940; AA078554.1; -.
DR GO; GO:0019866; C:inner membrane; IEA.
DR GO; GO:0016758; F:transferase activity, transferring hexosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0030259; P:lipid glycosylation; IEA.
DR GO; GO:004276; Glyco trans. 28.
DR InterPro; IPR004276; Glyco trans. 28.
DR InterPro; IPR007235; Glyco trans. 28.
DR InterPro; IPR006009; Murg.
DR Pfam; PF03033; Glyco trans. 28; 1.
DR Pfam; PF04101; Glyco trans. 28 C; 1.
DR TIGRfam; TIGR01133; murG; 1.
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 372 AA; 40242 MW; 2229D02DF85921A CRC64;

Query Match 25.3%; Score 476; DB 16; Length 372;
Best Local Similarity 33.6%; Pred. No. 8.3e-26;
Matches 124; Conservative 78; Mismatches 143; Indels 24; Gaps 11;

Qy 8 RLMWAGGTGGHVPGLAVAHLM--MACGVQVRLGTADRMADLVPKHGIEIDFIRISG 65
Db 6 RIISGGTGGHIFPAVSIANAIIELRPDAKILFVGAERMEMQVRPDAGYKIIGLPDAG 65
Qy 66 LRKGKIKALIAAPLRFNNAWRQARIMKAYKPDVVLGMGVGPGGLAANSLGIPVVLH 125
Db 66 PDRKHLKMNVSVLKLAESQWKAESLIKNRFPQVAVGVGGYASGPTLKTAGMVGPTLIQ 125
Qy 126 FONGIAGTNTKWLARIATKVMQAEPGA---PPNAEVV--GNPVTDLVLALEPQORLAGR 180
Db 126 EONSAYGVNTKLLAQKAKAICVAYDGMKEKFPADKIINTGNPVRQN-LTKDMPKGAALR 184
Qy 181 EGPVR-----VLVVGSGGARILNQTMPQVAUKL---GDSVIIHQSGKGSQQSVQAY 232
Db 185 SFNLPDQKTKLIVGGSLGARTINNTLTAALATIKENNDIQFIW-QTGKYYYPQVTEAVR 243
Qy 233 EAGQ-PQHKVTEITDDMAAAVAVADVVVCRSGALTVEIAAAGLPALFVFPQH-KDRQY 290
Db 244 AAGELPNLYTDFIKDMAAAVAAASDLVISRAGAGSISEFCILLHKFVLPVSPNVAEDHQT 303
Qy 291 NNALPLEKAGAAKITEIQPQLS---VDVANTLAGWSRETLTMAERARAASIPDAT 347
```

DB 304 KNALALVDKQARIIYKQSEAEAKLMDVALNTVA--DORLKELSENIAKLALPDSABIIA 361
QY 348 NEVSRAVA 356
DB 362 QEVKLAEA 370

RESULT 16
Q820F6 PRELIMINARY; PRT; 363 AA.
AC Q820F6;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Putative
DE UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)pyrophosphoryl-
DE undecaprenol N-acetylglucosamine transferase.
GN MURG OR SAV6122.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis";
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL; AF050545; BAC73833.1; -;
DR GO; GO:0019866; C:inner membrane; IEA.
DR GO; GO:0016758; P:transferase activity; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0030255; P:lipid glycosylation; IEA.
DR GO; GO:0019277; P:UDP-N-acetylgalactosamine biosynthesis; IEA.
DR InterPro; IPR004276; Glyco_tran_28.
DR InterPro; IPR007235; Glyco_tran_28_C.
DR InterPro; IPR006009; MurG.
DR Pfam; PF03033; Glyco_transf_28; 1.
DR Pfam; PF04101; Glyco_tran_28_C; 1.
DR TIGRfam; TIGR01133; murG; 1.
KW Transferase; Complete proteome.
SQ SEQUENCE 363 AA; 38524 MW; B085F2493277597C CRC64;

Query Match 24.5%; Score 461.5; DB 16; Length 363;
Best Local Similarity 34.9%; Pred. No. 8.6e-25;
Matches 130; Conservative 62; Mismatches 149; Indels 31; Gaps 10;

QY 9 LMVAGGTGGHVPGLVAHHLMAQGV--RWLGTADRMADLVPKHGIEIDIRISGL 66
DB 3 VVLAGGTGAGHIEPALADALRRDPTVGITGTERGLETRLPVPERGYDLALIPAVPL 62
QY 67 RKGKIKALIAAPLRFNWRQARATMKAVKPDVLMGMYGSGPGLAASLGI;PVVLHE 126
DB 63 PRKPTPELITVPCRIRGCIKAEOILETKADAVVGFGYVALGYLAANKELGVFIHIE 122
QY 127 QNGIAGLTNKLARIATKVMQEPGA--FNFVWGNVVRTDVALP-----LPQORLAGRE 181
DB 123 ANARPGLANKIGSRVAAQAVVSTPDSKLRGARYIGIPLRRSTIATIDRAAVRFEARAAFL 182

QY 182 GP--VRLVVGSGQARILNQTMPQVAAKLGDSVI-IWHQSGKSGSQSVQVQYAEAGQ 238
DB 183 DPNLPTLLVSGSGQARRLNEVYVQVAPYLOAGIQLH--AVGPKNEPQVHMPGMP 240
QY 239 HKVTEFIDMAAAYAWADVWCRSGLATVSEIAAAGLPALFVPFQHKDQOYVNALPLEK 298
DB 241 YIPVPYVDRMDLAYAADMMLCRAGMTVAELSAVGLPAAYVPLPIGNGEQRINACPVVK 300
QY 299 AGAAKIIEQPQSVDAVANTLAGWSRETL-----TMAERARAASIPDATERVANE-- 349
DB 301 AGGGLLVDDAELTPE-----WVQGNVLPLVADPHRLYENSRAS--EFGRDADLL 350
QY 350 VSRVABALEHHH 361
DB 351 VGMVYERIAARH 362

RESULT 17
Q8GBAO PRELIMINARY; PRT; 379 AA.
AC Q8GBAO;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl-(Pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.-) (Fragment).
OS Helicobacillus mobilis.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Helicobacteriaceae;
OC Helicobacillus.
OX NCBI_TaxID=28064;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22337798; PubMed=12446909;
RA Raymond J., Zhaxybayeva O., Gogarten J.P., Gerdes S.Y.,
RA Blankenship R.E.;
RT "Whole-genome analysis of photosynthetic prokaryotes";
RL Science 298:1616-1620(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Liolios K.G., Chu L., Ostrovskaya O., Mendybaeva N., Koukharenko V.,
RA Gerdes S., Kyrpides N., Overbeek R.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY142807; AAN87411.1; -;
DR GO; GO:0019866; C:inner membrane; IEA.
DR GO; GO:0016758; P:transferase activity; transferring hexosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0030259; P:lipid glycosylation; IEA.
DR GO; GO:0019277; P:UDP-N-acetylgalactosamine biosynthesis; IEA.
DR InterPro; IPR004276; Glyco_tran_28.
DR InterPro; IPR007235; Glyco_tran_28_C.
DR InterPro; IPR006009; MurG.
DR Pfam; PF03033; Glyco_transf_28; 1.
DR Pfam; PF04101; Glyco_tran_28_C; 1.
DR TIGRfam; TIGR01133; murG; 1.
KW Transferase; Glycosyltransferase.
FT NON_TER 379 379
SQ SEQUENCE 379 AA; 41095 MW; 4C2EB832148045EA CRC64;

Query Match 23.8%; Score 448.5; DB 2; Length 379;
Best Local Similarity 32.8%; Pred. No. 7.7e-24;
Matches 126; Conservative 71; Mismatches 138; Indels 49; Gaps 14;

QY 7 KRLMVAGGTGGHVPGLVAHHLMAQ--GWQVRLGTADRMADLVPKHGIEIDIRIS 64
DB 9 RKFVLGGGTGGHVIYALAIARGLQERFPFGCSIEYIGGRGLENTIVPEGPPLTKVHCR 68
QY 65 GL--RGKIKALIAAPLRFNWR--QARAIKAVKPDVLMGMYGSGPGLAASLGI 120
DB 69 GLERGLSLKMLA---IGTGRGLEALIFLRKLKPDVIGTGGFVAFVVMVATLLGI 124
QY 121 PVVLHEQNGIAGLTNKLARIATKVMQEPGAPNAE-----VVGNVVRTDVALP 171

```
Db 125 PALIHEQNAVGVNTRILAPRVGSMVLTTPFEACKRLKANDVIVTGLVPRPSILS-- 178
Qy 172 LPQORLAGR-----EGPVRLVVGSGQARILNOTMPOVAAKLG--DSVLIWHOSGKG 222
Db 179 --ASREGRKFPQIPPAQV--LLVVGSGRGAKLNEAMAPLARNLAGQRYQVHLVHTGSE 235
Qy 223 SQSQVEQAYAEAGOPQ-----KVTEFIDDMAAAYAWADVVCVRSALTIVSEIAAAGLPA 277
Db 236 NYDETRLLYEDAGIALEKCGNIKLLPYLDRMDVALAASDLVCVRAGAATFSEITARGLAS 295
Qy 278 LPVDFQK-KDQOQVWNLPLEKAGAAKIIIEQPOLSS-----VDVANTLAGRSRETLITMAE 332
Db 296 ILIPYAAENHOEANARSLESGATKVIDLRELTSQLQOEMVPHLLS--HROOVEMAAQ 353
Qy 333 RARAASIPDATERVANEVSRVARA 356
Db 354 AARSAGRPEALAHILSIERWKS 377

RESULT 18
Q8CV50 PRELIMINARY; PRT; 393 AA.
AC Q8CV50;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DE UDP-N-acetylglucosamine--N-acetylmutamyl- (pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase.
GN MURG OR BL1323.
OS Bifidobacterium longum.
OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacteriales;
OC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCC 2705;
EX MEDLINE=22294977; PubMed=12381787;
RA Schell M.A., Karmirantzou M., Snel B., Vilanova D., Berger B.,
RA Fessi G., Zwahlen M.-C., Desiere F., Bork P., Delley M.,
RA Fridmore R.D., Arigoni F.;
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
RT to the human gastrointestinal tract."
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427(2002).
DR EMBL; AE014760; AAN25123.1; -.
DR GO; GO:0019866; C:inner membrane; IEA.
DR GO; GO:0016758; F:transferase activity, transferring hexosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0030259; P:lipid glycosylation; IEA.
DR GO; GO:0019277; P:UDP-N-acetylglucosamine biosynthesis; IEA.
DR InterPro; IPR004276; Glyco_trans_28.
DR InterPro; IPR007235; Glyco_tran_28_C.
DR Pfam; PF03033; Glyco_transf_28; 1.
DR Pfam; PF04101; Glyco_tran_28_C; 1.
DR TIGRfams; TIGR01133; murG; 1.
KW Complete proteome.
SQ SEQUENCE 393 AA; 41511 MW; DAF16113F5923AAD CRC64;

Query Match 22.9%; Score 431.5; DB 16; Length 393;
Best Local Similarity 32.9%; Pred. No. 1.3e-22;
Matches 129; Conservative 60; Mismatches 156; Indels 47; Gaps 8;

Qy 5 QGKRLVMA--GGTGGHVPFGLAVAHHL--MAQGVQVRLGTADRMADLVPKHGIEDFI 61
Db 3 QGTPIHLVAGGTAGHVNPLAVAGAIRDIEPTAQTVIGTAGVLEKDLVEAGVELDTI 62
Qy 62 RISGLRGKIGKIALIAPRIFNWQRAIMKAYKPDVVLGNGGVVSGFGLAWSIGIP 121
Db 63 EKVPFPRRLNLYRFPKAKKETAKVRISILETRHADVVAGGVYASAPVYATARKGIP 122
Qy 122 VVHONGAGTGNKWLARIATKV-----MQAEFGAPFNARVGNPVRTDVLALPLPQQ 175
```

```
Db 123 TAIHEQNARAGMANKLGARWADFIVTVEGTGLKPRAGADVVERGLPLRPAIASL---TK 179
Qy 176 RIAGREGPVR-----VLVVGSGQARILNOTMPOVAAKLGDSVLIWHOSG 220
Db 180 RIGDDRAAVRESAAQLGVDNRLPLVLVTGSLGAQSLNRAIASSAADLLAHAQIIHLTG 239
Qy 221 KGSQSQVEQAY-----EAGOPQHKVTFIDDMAAAYAWADVVCVRSALTIVSE 266
Db 240 RGLKISEVRELVTASAGADVLTGIGPESAGQDHYHTAEYLERIDILAFACADLVICRAGKS 299
Qy 267 VSEIAAAGLPAIFVPFQHKDQOQVWNLPLEKAGAAKIIIEQ-----POLSDVAVANTLAGW 322
Db 300 VSELAALGLPLIYVPLPIGNGEQRFNAEPVNVNAGGLLVADKDLTPQWVHEHVPDLA-- 357
Qy 323 SRETLTWAERARAASIPDATERVANEVSRVA 354
Db 358 DHERLAEFGRKAWEXGIRNAAEIMARHVLQLA 389

RESULT 19
Q83HK1 PRELIMINARY; PRT; 356 AA.
AC Q83HK1;
DT 01-JUN-2003 (TREMELrel. 24, Created)
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)
DE UDP-N-acetylglucosamine--N-acetylmuramyl- (pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase (EC 2.4.1.1).
GN MURG OR TW542.
OS Tropheryma whipplei (strain TW08/27) (Whipple's bacillus).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococineae; Cellulomonadaceae; Tropheryma.
OX NCBI_TaxID=218496;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22495039; PubMed=12606174;
RA Bentley S.D., Maiwald M., Murphy L.D., Pallen M.J., Yeats C.A.,
RA Dover L.G., Norbertczak H.T., Besta G.S., Quail M.A., Harris D.E.,
RA von Herbay A., Goble A., Rutter S., Squares R., Squares S.,
RA Barrell B.G., Parkhill J., Rellman D.A.;
RT 'Sequencing and analysis of the genome of the Whipple's disease
RT bacterium tropheryma whipplei.'
RL Lancet 361:637-644(2003).
DR EMBL; BX251411; CAD67208.1; -.
DR GO; GO:0019866; C:inner membrane; IEA.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0016758; F:transferase activity, transferring hexosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0030259; P:lipid glycosylation; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR GO; GO:0019277; P:UDP-N-acetylglucosamine biosynthesis; IEA.
DR InterPro; IPR004276; Glyco_trans_28.
DR InterPro; IPR007235; Glyco_tran_28_C.
DR InterPro; IPR000524; HTH_Gntr.
DR InterPro; IPR006009; MurG.
DR Pfam; PF03033; Glyco_transf_28; 1.
DR Pfam; PF04101; Glyco_tran_28_C; 1.
DR TIGRfams; TIGR01133; murG; 1.
DR PROSITE; PS00043; HTH GNTR FAMILY; 1.
DR Transferase; Glycosyltransferase; Complete proteome.
KW Transferase.
SQ SEQUENCE 356 AA; 38420 MW; B555608CB471ED9D CRC64;

Query Match 22.9%; Score 430; DB 16; Length 356;
Best Local Similarity 33.1%; Pred. No. 1.5e-22;
Matches 121; Conservative 63; Mismatches 150; Indels 32; Gaps 10;

Qy 8 RLNVMAAGTGGHVPFGLAVAHHLMAQGVQVRLGTADRMADLVPKHGIEDFIRISGI- 66
Db 3 RIILAGGTAGHVNPLALADLVKVGSHATFATGTSGLSEGLVFNNSG--IDFFIPLKLP 60
Qy 67 -RGKIGKIALIAPRIFNWQRAIMKAYKPDVVLGNGGVVSGFGLAWSIGIPVLH 125
```

Db 61 FPARTSRHILCFPFKFFSSVKLVRSILIEHKIQVVVGGYVAAPAYAAASLNIPYVH 120
Qy 126 EQNGIAGLTNKLARIATKVNQABPGAPFNAEVVGNPVRTDVLAL-----PLPQORLAGRE 181
Db 121 ESNARPGLANLLAAHFAKCVGISVIGALPCGKLVGTPIRRDLTAAASFDFVLAKEKGLD 180
Qy 182 GPVR--VLVVGSGOGARILNQTW---POVAAKLGDSVLIW---HOSGKGSQOSVEQAYA 232
Db 181 -PVKLLLVFGSGGSAKINNHRAALPRVLKLCDEKNYLQVLIHITGYGDS-----I 232
Qy 233 EAGOPQHKVTEFIDDDMAAAYADVVVCRSGALTVEIAAAGLPALFVPFHQKDRQOYWN 292
Db 233 DVNPHYSSVRVYDMSMGYALSAADLVVSRAGSSTVAELCTFGIPAIYIYPFPGNGEQRN 292
Qy 293 ALPLEKAGAAKIIQOPOLSVDAVANTLAGWSRETLTMAERAASIPDATERVANEVSR 352
Db 293 VSHME--SAARIIGENDLSQIRLEDELL-----ELMTDDERREMSIAAKRFAICNAQN 345
Qy 353 VARALE 358
Db 346 TASLIE 351

RESULT 20

Q820Y4 PRELIMINARY; PRT; 356 AA.
AC Q820Y4;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE UDP-N-acetylglucosamine-N-acetylmuramyl- (pentapeptide)
DE pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE (EC 2.4.1.-)
GN MURG OR TWT228.
OS Tropheryma whipplei (strain Twist) (Whipple's bacillus).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococineae; Cellulomonadaceae; Tropheryma.
CX NCBI_TaxID=203267;
RN (1)
SEQUENCE FROM N.A.
RA Racult D., Audic S., Robert C., Ogata H., Suhre K., Drancourt M.,
EA Claverie J.-M.,
RT "Tropheryma whipplei illustrates the diversity of gene loss patterns
in small genome bacterial pathogens."
RL Submitted (JUL-2002) to the ENBL/GenBank/DBJ databases.
RL EMBL; AE016851; AAC4325.1;
DR GO; GO:0019866; C:inner membrane; IEA.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0001678; F:transferase activity, transferring hexosyl . . . ; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0030259; P:lipid glycosylation; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR GO; GO:0019277; P:UDP-N-acetylglucosamine biosynthesis; IEA.
DR InterPro; IPR004276; Glyco_trans_28.
DR InterPro; IPR007235; Glyco_trans_28_C.
DR InterPro; IPR000524; HTH_GntR.
DR InterPro; IPR006009; MurG.
DR Pfam; PF03033; Glyco_transf_28; 1.
DR Pfam; PF04101; Glyco_trans_28_C; 1.
DR TIGRFAMs; TIGR01133; murG; 1.
DR PROSITE; PS00043; HTH_GNTR_FAMILY; 1.
KW Transferase; Glycosyltransferase; Complete proteome.
SQ SEQUENCE 356 AA; 38420 MW; B55608CB471ED9D CRC64;

Query Match 22.9%; Score 430; DB 16; Length 356;
Best Local Similarity 33.1%; Pred. No. 1.5e-22;
Matches 121; Conservative 63; Mismatches 150; Indels 32; Gaps 10;
Qy 8 RLNMVAGTGGVFFGLAVAHHLMAQGNQVRNLTGADRMEADLVPKGIEIDFIRISGL- 66
Db 3 RILLAGGTAGHNPILLADVLKVSCHATPALGTSIEGRVPSNG--IDFTTIPKLP 60

Search completed: June 7, 2004, 07:15:18
Job time : 51 secs

Qy 67 -RGKIKALIAAPLRIFNMFQARAIMKAYKPDVVLGNGGVSGPGLAAWSLGIPIVVLH 125
Db 61 FPARTSRHILCFPFKFFSSVKLVRSILIEHKIQVVVGGYVAAPAYAAASLNIPYVH 120
Qy 126 EQNGIAGLTNKLARIATKVNQABPGAPFNAEVVGNPVRTDVLAL-----PLPQORLAGRE 181
Db 121 ESNARPGLANLLAAHFAKCVGISVIGALPCGKLVGTPIRRDLTAAASFDFVLAKEKGLD 180
Qy 182 GPVR--VLVVGSGOGARILNQTW---POVAAKLGDSVLIW---HOSGKGSQOSVEQAYA 232
Db 181 -PVKLLLVFGSGGSAKINNHRAALPRVLKLCDEKNYLQVLIHITGYGDS-----I 232
Qy 233 EAGOPQHKVTEFIDDDMAAAYADVVVCRSGALTVEIAAAGLPALFVPFHQKDRQOYWN 292
Db 233 DVNPHYSSVRVYDMSMGYALSAADLVVSRAGSSTVAELCTFGIPAIYIYPFPGNGEQRN 292
Qy 293 ALPLEKAGAAKIIQOPOLSVDAVANTLAGWSRETLTMAERAASIPDATERVANEVSR 352
Db 293 VSHME--SAARIIGENDLSQIRLEDELL-----ELMTDDERREMSIAAKRFAICNAQN 345
Qy 353 VARALE 358
Db 346 TASLIE 351